



This work is protected by copyright and other intellectual property rights and duplication or sale of all or part is not permitted, except that material may be duplicated by you for research, private study, criticism/review or educational purposes. Electronic or print copies are for your own personal, non-commercial use and shall not be passed to any other individual. No quotation may be published without proper acknowledgement. For any other use, or to quote extensively from the work, permission must be obtained from the copyright holder/s.

SYNCHROTRON X-RADIATION PROTEIN CRYSTALLOGRAPHY AND
ITS USE IN THE CRYSTAL STRUCTURE DETERMINATION OF
PURINE NUCLEOSIDE PHOSPHORYLASE

Thesis submitted for the degree of Doctor of Philosophy
at the University of Keele, April 1987

Stephen Anthony Rule B.Sc. (Hons)

Department of Physics

DECLARATION

I hereby declare that, unless otherwise stated, the work presented herein is my own and has not been previously submitted for a degree at this or any other university.

A handwritten signature in dark ink, appearing to read 'Rule', is centered on the page.

STEPHEN ANTHONY RULE

Acknowledgements

I should like to express my thanks to Dr. John Hellivell: firstly, for stimulating an interest in protein crystallography during my undergraduate days in physics, and secondly, as my supervisor, for his guidance, encouragement, and enthusiasm.

I am grateful to Professor Watson Fuller and the Department of Physics for the provision of a grant and the use of facilities; to Dr. Trevor Greenbough for sharing his understanding and experience; and to my contemporaries for our many useful discussions.

The SERC and all the staff of the Daresbury Laboratory are thanked for providing synchrotron radiation beam time and support.

I was fortunate enough to spend summer '82 and spring '84 in Birmingham, Alabama, in the laboratory of Professor Charlie Bugg, to whom I am indebted for making those visits possible. I owe special thanks to Drs. Steve Halick, Jack Sack, Alan Zell (sadly no longer with us), and many other friends, for helping to make those times both educational and productive in the lab., and a pleasure out.

For typing this thesis, in what amounts to a foreign language, I am extremely grateful to Caroline Baggeley; her patience is clearly unlimited.

Finally, words cannot adequately express my gratitude to my wife, Trina, for her invaluable support, both academic and emotional.

Abstract

The availability of intense, tunable Synchrotron X-radiation (SR) has, in recent years, revolutionised certain aspects of macromolecular crystallography. Data collection times have been dramatically reduced, the range of suitable samples has been widened, and techniques such as optimised anomalous scattering and Laue diffraction have become feasible.

An understanding of the characteristics of the SR beam and its interaction with single crystal monochromator systems leads to the possibility of a polychromatic diffraction experiment in which each ray's energy is correlated with its direction within the beam. Furthermore, this correlation is preserved in the reflections from a crystal sample, thereby allowing the wavelength dependent effects of an anomalous scatterer to be recorded simultaneously and yet remain resolved spatially. The potential of this approach is demonstrated by an experiment in which a small Re complex was exposed to a polychromatic beam centred on the Re L_{III} absorption edge wavelength.

Purine nucleoside phosphorylase (PNP) is an important enzyme in purine catabolism and salvage. PNP deficiency is associated with selective T-cell immunodeficiency and, consequently, an effective PNP inhibitor might prove to be useful in treating various conditions including T-cell leukemias. With a view towards rational drug design, the crystal structure of PNP was initially determined at 6.0 Å

resolution using diffractometer data. However, the crystals diffract weakly and data to high resolution from native and derivative crystals were therefore collected at the Daresbury SRS.

Phases were calculated by the method of isomorphous replacement and the final electron density map at 3.16 Å resolution is of high quality, enabling the entire polypeptide chain to be followed. The structure consists principally of a six stranded β sheet flanked by several α -helices. The region of the active centre is seen to be accessible to the large solvent channels running throughout the crystal lattice. Data from a variety of PNP - substrate and - substrate analog complexes have also been measured and should allow a detailed characterisation of the structural features important for the enzyme's mechanism.

CONTENTS

	Page
1. Introduction	1
1.1 Fundamentals of X-ray crystallography	3
1.2 Isomorphous replacement and anomalous scattering	10
1.3 Characteristics and applications of synchrotron X-radiation	25
1.4 Drug design and crystallography	42
2. The SR polychromatic profile method in crystallography	49
2.1 Theory	50
2.2 Experimental	63
3. Properties and significance of Purine Nucleoside Phosphorylase	72
3.1 The function of PNP	73
3.2 Significance of PNP in chemotherapy	75
3.3 Properties of PNP	79
3.4 Inhibitors of PNP	85

	Page
4. Data collection and reduction	90
4.1 Preliminary crystallographic studies	90
4.2 Heavy atom derivative search	93
4.3 Data collection at the SRS	96
4.4 Data reduction	102
4.5 Analysis of the derivatives	111
4.6 Data collection on the wiggler beam line	113
5. The 6 Å structure of PNP	115
5.1 Data collection and calculation of phases	115
5.2 The electron density map at 6.0 Å resolution	118
5.3 Substrate binding studies	120
6. The 3.16 Å structure of PN?	123
6.1 Heavy atom parameters	123
6.2 The absolute configuration	125
6.3 The calculation of phases	127
6.4 The electron density map	128
6.5 Substrate binding studies	135
7. Conclusion	137
References	
Appendices	

CHAPTER ONE

INTRODUCTION

Since Kendrew et.al. (1958) revealed the three dimensional structure of myoglobin, many other proteins have been successfully studied by the methods of X-ray crystallography, thereby enhancing the understanding at the atomic and molecular level, of enzyme catalysis, electron transfer, oxygen storage and transport, hormone function, the immune system, protein synthesis, and protein evolution. However, protein crystallography remains highly labour intensive, and, although advances in computer power and computer graphics have greatly facilitated the analysis of macromolecular structures, there is enormous interest in the development of new techniques which may lead to economies in the time and effort invested between the initiation of an investigation and the unveiling of significant biological information. Thus, the availability of intense synchrotron X-radiation has stimulated consideration of the methods of crystal data collection and of phasing the individual reflections. In addition to this it promises the potential to tackle previously non-viable studies of small or very poorly diffracting samples, and to obtain time resolved information.

This thesis is concerned, firstly, with the synchrotron radiation data collection from weakly scattering crystals,

and the subsequent structural analysis, of the enzyme purine nucleoside phosphorylase, and secondly, with the initial development of a new polychromatic anomalous dispersion method. This chapter, therefore, presents brief reviews of the basic methods of protein crystallography and of the relevant characteristics and applications of synchrotron X-radiation. A later chapter (§3) deals with the properties of purine nucleoside phosphorylase.

Through the achievements of molecular biology has come an awareness of the complexity of biological systems; not only are there millions of proteins worthy of investigation, but, in addition, each protein has a variety of possible conformational states depending on its local environment. Thus, increasing emphasis is placed on the need to select those studies in which structural knowledge, once acquired, can be applied to the greatest effect, or relates to key processes in biology. There is, therefore, at least in the present socio-economic climate, a demand for some reconciliation of the interests of pure academic research and the perceived needs of industry. It is in this context that the fields of protein engineering and rational drug design have been placed. It is now possible, through the technique of site-directed mutagenesis, to make specific alterations in the amino acid sequence of a protein, thereby modifying its behaviour in some way (reviewed by Perutz (1985)) and producing, for example, a more efficient industrial catalyst. An understanding of the three dimensional structure of the

catalytic site is required to pinpoint those residues suitable for modification.

The prime reason for initiating a crystallographic investigation of purine nucleoside phosphorylase is that improved inhibitors of the enzyme have real potential, both directly and indirectly, in the treatment of certain diseases. Knowledge of the active site structure, of the catalytic mechanism, and of the interaction with known inhibitors should provide the means with which to develop new inhibitors leading eventually to clinically useful drugs. The final section of this introductory chapter outlines some of the concepts involved in a rational approach to drug design.

1.1 Fundamentals of X-ray crystallography

There are many excellent texts covering the relevant diffraction theory (e.g. Woolfson (1970)) and, in particular, Flinn and Johnson (1976) have reviewed all aspects of protein crystallography.

1.1.1 Diffraction theory and the phase problem

When a beam of X-rays impinges upon matter, the individual contributions of elastically scattering electrons combine to produce a diffraction pattern. Consider first a single atom where the electron density at a distance \underline{r} from its centre is $\rho(\underline{r})$. If the incident wave vector is \underline{a}_0 and the scattered wave vector is \underline{a}' , then the

total wave scattered by a volume element dv at a position \underline{r} , relative to the wave scattered by a unit electron at the origin, is

$$\rho(\underline{r}).\exp(2\pi i \underline{r} \cdot \underline{S}) dv ,$$

where $\underline{S} = \underline{s}_0 - \underline{s}'$ (figure 1.1). Hence, the total wave scattered by the atom is,

$$\underline{f}(\underline{S}) = \int_{\text{Vol. atom}} \rho(\underline{r}).\exp(2\pi i \underline{r} \cdot \underline{S}) dv ,$$

and this represents the 'atomic scattering factor'. Some typical atomic scattering curves are shown in figure 1.2.

For the j th atom, at a distance of \underline{r}_j from the new origin, in a unit cell of N atoms

$$\underline{f}_j = \int \rho(r).\exp(2\pi i (\underline{r}_j + \underline{r}) \cdot \underline{S}) dv$$

$$= f_j.\exp(2\pi i \underline{r}_j \cdot \underline{S})$$

where f_j replaces $f(\underline{S})$. The total wave scattered by all the atoms is then,

$$\underline{G}(\underline{S}) = \sum_{j=1}^N f_j.\exp(2\pi i \underline{r}_j \cdot \underline{S})$$

and this represents the 'molecular transform'. However, since the molecular transform turns out to be very small, a diffraction pattern will only be observable when constructive interference occurs, i.e. when the phase

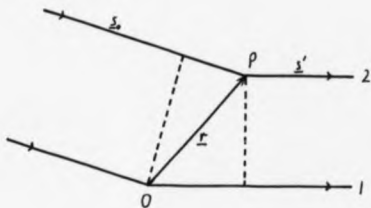


Figure 1.1 Scattering from a point P relative to an origin O.
The path difference between ray 2 and ray 1 is $\lambda \underline{s} \cdot \underline{s}_s - \lambda \underline{s} \cdot \underline{s}'$.

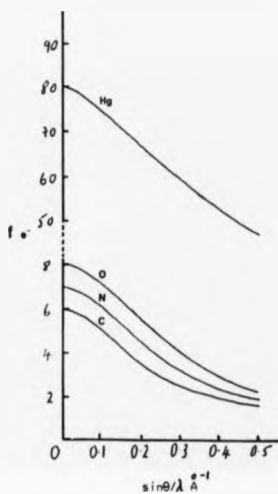


Figure 1.2 Atomic scattering factor curves for C, N, O and Hg, a common heavy atom substituent in protein crystallography.

difference between the waves scattered by successive unit cells (defined by the vectors \underline{a} , \underline{b} and \underline{c}) within the crystal is an integral multiple of 2π , in which case the total wave scattered becomes

$$F(\underline{S}) = T \cdot G(\underline{S}),$$

where T is the total number of unit cells.

The condition for observable diffraction is embodied in the Laue equations:

$$\underline{a} \cdot \underline{S} = h, \underline{b} \cdot \underline{S} = k, \underline{c} \cdot \underline{S} = l,$$

where h , k and l are integers. If the general position vector \underline{r}_j is now expressed in fractional cell coordinates, i.e.

$$\underline{r}_j = \underline{a}x_j + \underline{b}y_j + \underline{c}z_j,$$

then

$$\underline{r}_j \cdot \underline{S} = hx_j + ky_j + lz_j,$$

and

$$F(hkl) = F(\underline{S}) = \sum_{j=1}^N f_j \cdot \exp 2\pi i (hx_j + ky_j + lz_j),$$

where the constant of proportionality has been neglected. This last expression is the 'structure factor equation', and represents the molecular transform sampled at the 'reciprocal lattice' points, hkl . The reciprocal lattice

is itself the Fourier transform of the real crystal lattice, and each reciprocal lattice point corresponds to a set of reflecting planes within the crystal. The inverse transform of the structure factor equation gives the electron density equation, namely

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} \underline{F}(hkl) \cdot \exp - 2\pi i(hx + ky + lz) ,$$

where V is the volume of the unit cell and the summation is over the entire volume of reciprocal space.

The conditions for diffraction can be conveniently expressed via the geometric construction of Ewald (1921), shown in figure 1.3. The locus of the scattered wave vector, \underline{s}' , is a sphere of unit radius centred at the crystal. If the reciprocal lattice is centred at O where the unreflected beam meets the sphere, then diffraction can occur in the direction of \underline{s}' when the point B coincides with a reciprocal lattice point, i.e. when a reciprocal lattice point intersects the sphere of reflection.

To reconstruct the contents of the unit cell, all that is needed are the structure factors, $\underline{F}(hkl)$. In general, however, $\underline{F}(hkl)$ is a complex quantity:

$$\underline{F}(hkl) = F(hkl) \exp i\alpha(hkl),$$

where $F(hkl)$ is the amplitude, and $\alpha(hkl)$ the phase, of the structure factor. In recording a diffraction pattern, only the intensities ($= |\underline{F}(hkl)|^2$) of the diffracted rays

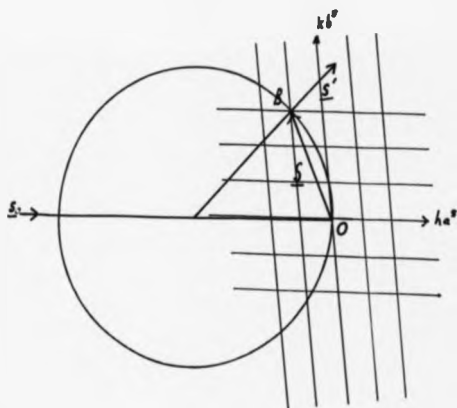


Figure 1.3 The Ewald construction representing diffraction in reciprocal space. The reciprocal unit cell is defined by the vectors $\underline{a}^*, \underline{b}^*, \underline{c}^*$, which are inversely related to $\underline{a}, \underline{b}, \underline{c}$.

can be measured and the phase information is lost. Herein lies the central problem of structure determination by X-ray diffraction: to obtain by analytical and/or experimental means some estimate of the structure factor phases.

For the determination of small (<100 atoms) molecular structures 'direct' methods (see Woolfson (1971) for a review), using mathematical relationships between structure amplitudes, can generally be used, but these methods do not lend themselves to the solution of comparatively large protein structures. The molecular replacement method (Rossmann (1972)) has been very successful in solving proteins for which there is a highly homologous protein of known structure available. However, nearly all the ab initio protein structures reported to date have relied upon the methods of isomorphous replacement (Green et.al. (1954), Blow and Crick (1959)), either with or without the additional information obtained from anomalous dispersion (North (1965), Matthews (1966)).

1.1.2 Measurement of diffraction intensities

Three major problems are encountered in the collection of X-ray data from protein crystals:

- a) the inherent weakness of the individual diffracted beams,
- b) the large number of reflections to be measured,
- c) the damage inflicted on the sample by the absorption of X-ray photons.

The first aspect of data collection that needs to be considered is that of the radiation source. In a conventional X-ray generator, an accelerated beam (typically 40kV) of electrons from a heated filament, bombard a metal target and stimulate the emission of a line spectrum characteristic of the target (figure 1.4). The target can be stationary and enclosed in a sealed tube, or, alternatively, it can be rotated to spread the heat load. A typical sealed tube may operate at 1.2kW for 1000 hours whilst a rotating anode generator can deliver up to twice that power with an electron focus eight times smaller, albeit with some reduction in reliability. Tunability of the wavelength is limited to the accessible K_{α} and K_{β} emission lines of a few suitable target materials (e.g. Cu K_{α} 1.54 Å, Co K_{α} 1.79 Å, Mo K_{α} 0.71 Å). The spectral spread of the X-ray beam can be reduced by using a Ni filter in the case of Cu radiation (see figure 1.4), or by reflecting the beam from a single crystal of graphite.

The limited intensity of monochromatic X-rays available from a conventional source has, in the past, severely hampered many macromolecular structural investigations. Recording a high resolution ($d_{\min} < 2.5$ Å) data set can often span many months due to a slow rate of data collection and the large volume of data required. Indeed, it is often the case that such a data collection is impossible because each sample effectively 'dies' from radiation damage before enough high resolution terms can

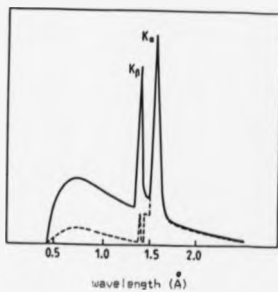


Figure 1.4 The spectrum emitted by a conventional X-ray source fitted with a Cu target. The dotted curve represents the effect of a Ni filter.

be recorded with suitable accuracy.

Synchrotron radiation (SR) was long regarded as an unwanted energy loss mechanism in the particle accelerators developed for high energy physics research into sub-atomic structure, and so the early experiments exploiting the radiation were conducted under far from ideal conditions. Since that time, SR has become well established at the forefront of modern research areas in physics, chemistry and biology, to the extent that synchrotrons and storage rings are now being designed solely for the production of intense, broad-band radiation. The Daresbury Synchrotron Radiation Source (SRS), the first dedicated source when commissioned in 1981, is now capable of delivering a highly collimated, tunable, monochromatic X-ray beam which, at the sample, is up to 1000 times more intense than a typical rotating anode source. The use of such intense sources has led, not only to the greatly improved data collection rates that were expected, but also to an increase in typical crystal lifetimes, thereby enabling previously impractical studies to be initiated.

A full introduction to the characteristics of SR, and their application to macromolecular crystallography, is presented in §1.3.

The second aspect of data collection that needs to be considered is the method itself, i.e. the diffraction geometry and the detector used. Computer controlled automatic diffractometers were developed to facilitate the collection of large numbers of reflections (Arndt and

Willis (1966)). On a four circle diffractometer, reflections are measured one at a time, generally by step scanning the crystal through the reflecting position and recording the diffracted intensity with a stationary scintillation counter. Such methods are suitable for collecting high resolution data from low molecular weight proteins, and low resolution data from high molecular weight proteins. The use of a diffractometer to measure low resolution protein data is described briefly in §5.1. For high resolution data collection from high molecular weight proteins the screenless rotation method (Arndt et.al. (1973), Arndt and Wonacott (1977)), where all the reflections (often more than 250) which are in the reflecting position at the same time, are recorded simultaneously by an area sensitive detector, offers a considerable gain in efficiency over diffractometer methods. The rotation, or oscillation, method was used in the present work and is described in §4.3.

Until recently, photographic film was the only suitable area detector available, however, multi-wire proportional chambers (MWPC) have been used to solve protein structures (Matthews et.al. (1977), Xuong et.al. (1978)). In addition, TV detectors are under development and seem likely to be highly attractive for use in protein data collection. A comparison of these various detectors, in the context of SR, is made in §1.3.

1.2 Isomorphous replacement and anomalous scattering

The method of isomorphous replacement relies upon

the availability of suitable heavy atom derivatives of the native protein. In some cases this can be done by replacing a metal atom already present with a heavier one, (e.g. zinc by lead in the case of insulin). Generally, however, derivatives are prepared by heavy atom addition rather than replacement. Protein crystals offer the opportunity for heavy atom reagents to diffuse through their large solvent channels and bind to the exposed surface of the protein without causing conformational changes or disrupting the crystal lattice.

Provided that the heavy atom is sufficiently electron dense, and the degree of substitution sufficiently high, observable changes in the diffraction pattern can be produced. The phase problem may then be solved vectorially, once the heavy atom positions within the unit cell have been determined. The use of one isomorphous derivative (i.e. single isomorphous replacement or SIR), leaves an ambiguity in the protein phase (figure 1.5), except for centric reflections where the phase is uniquely determined (figure 1.6). For acentric reflections, a second derivative is required to resolve the phase ambiguity, and, in practice, it is advisable to include information from more than two derivatives (multiple isomorphous replacement or MIR) since errors in the intensity measurements and the heavy atom parameters, mean that the phase circles do not intersect at one point.

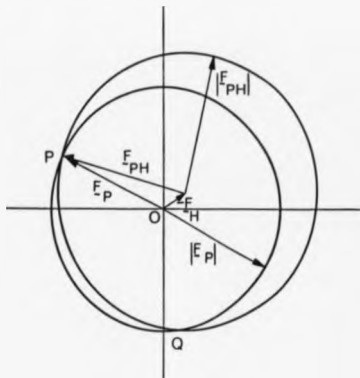
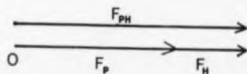
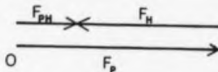


Figure 1.5 Single isomorphous replacement phase determination.

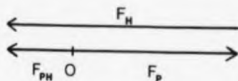
The magnitudes $|F_P|$ and $|F_{PH}|$ are known experimentally, and F_H , α_H are calculated from the known heavy atom structure. The vectors OP and OQ are the two possibilities for F_P .



(a)



(b)



(c)

Figure 1.6 Phase determination in the case of centric reflections.

a) $F_{PH} > F_P$ and $s(F_P) = s(F_H)$, where $s(F_P)$ is the sign of F_P .

b) $F_P > F_H$ and $s(F_P) = -s(F_H)$.

c) the 'crossover' case where $F_H > F_P$ and $s(F_P) = -s(F_H)$.

When $F_P = F_{PH}$ the phase remains ambiguous.

1.2.1 Anomalous scattering

In §1.1 it was assumed that the electrons of atoms in crystals scatter as if they were free electrons. This is, however, not true, particularly at wavelengths close to atomic X-ray absorption edges which correspond to the resonance frequencies of the inner K and L shell electrons. The atomic scattering factor has to be rewritten to incorporate these anomalous scattering effects:

$$f = f_0 + f'(\omega) + if''(\omega),$$

where f' and f'' are respectively the real (dispersion) and imaginary (absorption) components of the anomalous scattering, and ω is the frequency of the incident wave. Figure 1.7 shows examples of the variation of f' and f'' with wavelength.

For crystal structures containing only light atoms (C, N, O, H), f' and f'' are negligible at X-ray wavelengths, and the amplitudes of reflections hkl and $\bar{h}\bar{k}\bar{l}$ are equal. This is known as Friedel's law. If, however, an anomalous scatterer is introduced (i.e. a heavy atom), the absolute nature of f'' causes the breakdown of Friedel's law, as is shown in figure 1.8. The anomalous difference, δ , is defined as $F_{hkl} - F_{\bar{h}\bar{k}\bar{l}}$, and can be used in conjunction with the isomorphous difference to solve the phase problem, (single isomorphous replacement with anomalous scattering or SIRAS) (figure 1.9).

The changes in intensities due to anomalous scattering

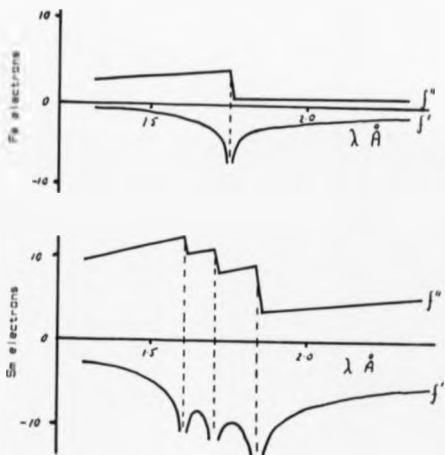


Figure 1.7 Idealised curves for the real (f') and imaginary (f'') terms of anomalous scattering due to the K edge of Fe and the L edge of Sm. (from Hoppe and Jakubovski (1975))

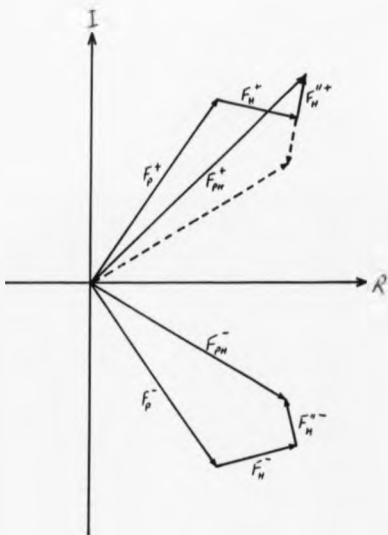


Figure 1.8 The imaginary component (I'') of anomalous scattering causes the breakdown of Friedel's law. The dashed vectors represent the inverse reflection mirrored across the real axis.

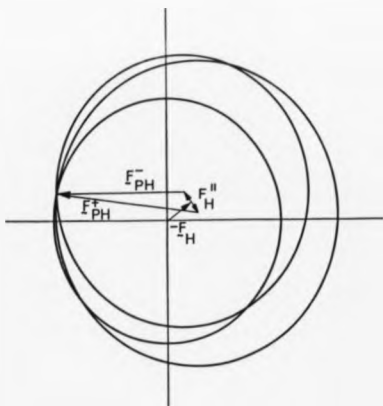


Figure 1.9 Single isomorphous replacement with anomalous scattering. In the absence of errors the anomalous information (as with a second derivative) uniquely resolves the protein phase.

will be less than for isomorphous replacement (see table 1.1) and, consequently, accuracy of measurement is crucial. However, anomalous differences are not subject to the systematic errors due to non-isomorphism, and they also become more significant at high resolution. This is because the anomalously scattering electrons are tightly bound close to the nucleus and so the scattering does not suffer the same intrinsic fall-off with $\sin\theta/\lambda$ as the normal scattering from outer electrons, although the fall-off due to thermal vibration and static crystal disorder occurs for all electrons.

A large majority of the protein crystal structures reported to date, have relied upon multiple isomorphous replacement with anomalous scattering, thus emphasizing the point that, because of errors in intensity measurement, it is desirable to include as many independent estimates as possible. Recently, however, anomalous differences alone have been used to solve protein structures. Smith et.al. (1983) used the anomalous scattering from the three di-iron centres in a trimeric haemerythrin to calculate protein phases to 5.5 Å resolution. Hendrickson and Teeter (1981) solved the crystal structure of crambin at 1.5 Å resolution, using only the anomalous scattering from six sulphur atoms, using a wavelength (1.54 Å) well removed from the sulphur K absorption edge (5.02 Å). In both cases, the phase ambiguity was resolved by biasing the phase probability distribution towards that of the partial

		N.W. in asymmetric unit			
		10000	30000	100000	300000
f^a	$1 \times I$	40%	23%	13%	7%
	$1 \times Hg$	60%	35%	19%	11%
	$2 \times Hg$	86%	49%	27%	15%
	$1 \times 5 e^-$	7.6%	4.4%	2.4%	1.4%
	$1 \times 10 e^-$	15%	9%	5%	3%
	$1 \times 20 e^-$	30%	18%	10%	6%

Table 1.1 Expected r.m.s. intensity changes due to heavy atom substituents and to anomalously scattering atoms. The fractional change is given by $(2N_h/N_p)^{1/2} \cdot f_h/f_p$ (Crick and Magdoff, (1956)), where N_h is the number of heavy (or anomalously scattering) atoms in the asymmetric unit, N_p is the number of non-hydrogen protein atoms, f_h is the heavy atom scattering factor at $\theta=0$ (or twice f^a in the anomalous case), and f_p is the mean protein atomic scattering factor (f_e^-).

structure (Fe or S).

1.2.2 Determination of heavy atom positions

To apply the method of isomorphous replacement, the sites of substitution of derivative heavy atoms must first be located, thereby enabling the heavy atom contribution to the structure factors to be calculated. The direct methods universally used in small molecule crystallography can be applied to this problem (Wilson (1974)) and have proved useful in certain studies (e.g. Steitz (1968), Neidle (1973), Navia and Sigler (1974)). Direct methods have not been used in the present work and so will not be discussed further. The methods which have been applied, namely difference Pattersons and difference Fourier's, along with the principles of heavy atom parameter refinement, are now briefly described.

1.2.2.1 The difference Patterson method

The Patterson function is a convolution function defined as:

$$P(uvw) = \int_{\text{cell vol}} \rho(xyz) \rho(x+u, y+v, z+w) dv,$$

which reduces to:

$$P(uvw) = \frac{1}{V} \sum |F(hkl)|^2 \cos 2\pi(hu + kv + lw).$$

and represents a map of inter-atomic vectors. The function

can be used for the ab initio solution of simple small molecule structures. If a structure has N atoms in the unit cell the Patterson map contains $N(N-1)$ vectors, excluding N self vectors superimposed at the origin of Patterson space. Hence, a protein Patterson map is completely uninterpretable despite containing structural information. However, a synthesis with coefficients $|\underline{E}_H|^2 = (|\underline{E}_{PH}| - |\underline{E}_P|)^2$, contains peaks corresponding to the vectors between heavy atoms. This is the isomorphous difference Patterson and, except in adverse circumstances, can be deconvoluted to reveal the heavy atom partial structure.

The above expression for \underline{E}_H is exact only for centric reflections (except for the relatively unlikely crossover case where \underline{F}_P and \underline{F}_{PH} are both small and \underline{E}_H may equal $\underline{E}_P + \underline{E}_{PH}$); for acentric reflections it is less accurate unless \underline{E}_P and \underline{E}_{PH} happen to be colinear. The approximation introduces noise into the map from protein-heavy atom and protein-protein vectors. Problems of poor signal to noise generally only arise when there are multiple heavy atom sites of low occupancy.

1.2.2.2 The difference Fourier method

Once estimates of the protein phases are available from other derivatives, a 'cross phase' difference Fourier map may reveal heavy atom sites. A Fourier synthesis using the coefficients $\underline{F}_{PH} \exp(i\alpha_{PH})$, will give the protein

plus heavy atom electron density. In general, $\alpha_{PH} \approx \alpha_P$ and so coefficients $F_{PH} \exp(i\alpha_P)$ could be used to give similar electron density. The difference in electron density between the derivative and the native protein, i.e. the heavy atom electron density, is therefore given by a Fourier synthesis with coefficients $(F_{PH} - F_P)\exp(i\alpha_P)$. These coefficients are usually weighted by some estimate of the reliability of the protein phase.

Since Fourier syntheses are dominated by the phases rather than the amplitudes (Srinivasan (1961)), it is important that the phases used be obtained from independent derivatives. The model used to calculate the phases, will, due to errors, tend to be reflected as 'ghost' peaks in the difference map, and these must be distinguished from the true peaks. Common sites between two derivatives indicated by difference Fourier maps should be treated with care.

Analysis of the assumption that $\alpha_{PH} = \alpha_P$ reveals three terms contributing to the electron density. The first accurately represents the heavy atom structure (where the atoms appear at half their true height), the second contributes random noise and the third, whose amplitude is proportional to F_H^2/F_P , represents a distorted image of the protein structure. Hence, the difference Fourier becomes less useful as the degree of heavy atom substitution increases.

The technique is particularly useful in studying

the binding of small molecules (e.g. substrates or inhibitors) to proteins, once the native protein structure has been solved. The mean change in F is small and the protein phases are well determined (indeed, calculated phases from a refined model can be used in place of the experimentally determined phases). The resulting map will also reveal conformational changes induced in the native protein structure on binding such a complex.

Anomalous differences (based on f' or f'') may also be employed in both difference Fouriers and difference Pattersons. Used alone in these syntheses, the positions of anomalously scattering atoms may be located or confirmed. When combined with isomorphous differences to obtain a better estimate of the heavy atom structure factors (see below) a significant improvement in the signal to noise ratio in both types of map can be achieved

1.2.2.3 Refinement of heavy atom parameters

To determine the phase angle of a native protein structure factor the triangulation of vectors, $E_{PH} = E_P + E_H$, needs to be accurately closed, and to achieve this, a well defined heavy atom vector is required. The heavy atom parameters (position, occupancy and thermal vibration) are, therefore, refined by standard least squares techniques against the observed isomorphous differences.

If protein phases calculated from an independent derivative are available, they may be used in a 'phased' refinement (Dickerson et.al. (1961), Muirhead et.al. (1967),

Blow and Matthews (1973)). The lack of closure, ϵ of the structure factor triangle can be defined as,

$$\epsilon = F_{PH}(OBS) - F_{PH}(CALC) ,$$

and

$$F_{PH}(CALC) = |\sum P(CALC) + \sum H(CALC)|.$$

The quantity minimized in the refinement is the sum over all reflections,

$$\sum W(F_{PH}(OBS) - F_{PH}(CALC))^2 ,$$

and the weight, W , is the inverse of the r.m.s. lack of closure over all the derivatives used in the calculation of the protein phase. Alternatively, the weight may be taken as m , the figure of merit of the phase calculation (see below), and data with a high value of m (say > 0.8), are included.

This refinement scheme depends on accurate measurement of the native structure factors, the quality of the protein phases used, and the lack of common sites between the derivatives used to calculate the phases and the derivative being refined. Despite these requirements, the method has been widely and successfully employed in protein structure determinations.

The second method of heavy atom refinement does not require any estimates of the protein phase, and can, therefore, be used in the absence of other derivatives.

It does, however, require an accurate estimate of the heavy atom structure amplitude from the observed diffraction intensities. For centric reflections, this presents no difficulty (except in the instances of possible crossover) since $F_H = F_{PH} - F_P$. Where crystals have sufficient data in centric zones, a centric only refinement will generally give perfectly adequate results.

For acentric reflections $|F_{PH} - F_P|$ will be a poor estimate of F_H , unless F_P and F_{PH} are nearly colinear. However, inclusion of anomalous scattering information allows the calculation of the heavy atom lower estimate, F_{HLE} (Harding (1962), Dodson and Vijayan (1971)), given by taking the negative sign in the expression,

$$F_H^2 = F_P^2 + F_{PH}^2 \pm 2F_P F_{PH} \left(1 - \frac{k^2}{2}\right)^{\frac{1}{2}}$$

where k is the ratio of the real and the imaginary components of the heavy atom scattering factor, F_H/F_H'' . Unreliable reflections, where the '+' sign in the above may apply, can be excluded. The quantity minimized in the least squares refinement is the residual,

$$R = \sum (F_{HLE} - F_{H(CALC)})^2$$

where $1/W$ is the variance of F_{HLE} .

Due to their relatively large errors, the anomalous differences (and, hence, F_{HLE}) will tend, on average, to be overestimated. To correct for this, Dodson et al. (1975) suggested subtracting from F_{HLE} a bias correction

$k^2\sigma^2/4$, where σ^2 is the variance of the anomalous difference. Alternatively, an empirical value of k , in ranges of $\sin^2\theta/\lambda^2$, and excluding centric reflections, can be used to reduce the bias, where k_{emp} is given by twice the r.m.s. isomorphous difference over the r.m.s. anomalous difference (Matthews (1966)).

These different refinement methods have each been successfully used in various protein structure determinations, and, where direct comparison is available, e.g. rubredoxin (Herriott et.al. (1970)) and insulin (Adams et.al. (1969)), are seen to yield results that are in good agreement.

1.2.3 Calculation of phases

In §1.2, it was shown that in the absence of errors, the phase of the native structure factor can be uniquely determined for the acentric reflections from two sources of independent information. In practice, the presence of errors will always lead to some uncertainty in the phase. Blow and Crick (1959) have shown that proper consideration of the errors leads to a phase probability, from which the 'best' protein phase, and a suitable weight for the electron density calculation, can be derived. They make the assumption that F_p is accurately determined and all experimental error (δ) is assigned to F_{PH} . The error due to non-isomorphism and to inaccuracies in the heavy atom model, (ϵ), is then assigned to $F_{H(CALC)}$. They then proceed to show that the error probability distribution can be represented by a Gaussian of breadth $\langle E \rangle = (\langle \epsilon \rangle^2 + \langle \delta \rangle^2)^{1/2}$, and that the

whole error can be regarded as residing in F_{PH} . If the 'lack of closure' of the phase triangle is defined as

$$\epsilon(\alpha) = |F_{PH}| - |E_P + E_H(\alpha)|,$$

then the probability that a phase angle, α , for the native structure factor is correct is given by,

$$P(\alpha) = \exp\left(-\frac{\epsilon(\alpha)^2}{2E^2}\right),$$

In practice, E is generally taken to be the r.m.s. lack of closure error in the appropriate region of reciprocal space (see §6.3).

The above formalism has also been extended to anomalous scattering information (North (1965), Matthews (1966)), where the lack of closure is defined by

$$\epsilon^*(\alpha) = \Delta_{obs} - \Delta_{calc}(\alpha)$$

The overall phase probability distribution combining information from several sources, is given by the product of the individual probabilities. Blow and Crick (1959) showed that the protein Fourier with the minimum r.m.s. error is that calculated with the coefficients $m_F \exp(i\alpha_B)$, where (m, α_B) are the polar coordinates of the centre of gravity of the phase probability distribution around a phase circle of unit radius. The components (m, α_B) can be obtained from

$$m \cos \alpha_B = \frac{\int_0^{2\pi} P(\alpha) \cos \alpha \, d\alpha}{\int_0^{2\pi} P(\alpha) \, d\alpha}$$

$$m \cdot \sin \alpha_B = \frac{\int_0^{2\pi} P(\alpha) \sin \alpha \, d\alpha}{\int_0^{2\pi} P(\alpha) \, d\alpha}$$

$$m = \frac{\int_0^{2\pi} P(\alpha) \exp(i\alpha) \, d\alpha}{\int_0^{2\pi} P(\alpha) \, d\alpha}$$

If the phase probability for a reflection is fairly uniform around the circle, the centroid of the distribution will be close to the origin and so m will be close to zero. Hence, m , the 'figure of merit', is a weighting function which reflects the sharpness of the phase determination.

Once phases have been calculated, all that remains is the straightforward computation of the Fourier series summation to give the electron density,

$$\rho(xyz) = \sum_{hkl} F_p \exp(i\alpha_p) \exp(-2\pi i(hx + ky + lz))$$

1.2.4 Interpretation of the electron density and structure refinement

A protein structure analysis generally proceeds in stages from low through to high resolution. Although the information contained in a low resolution (5 - 6 Å) electron density map is limited, it is still useful if only to assess the progress to that point. It will not normally be possible to trace the path of the polypeptide chain,

although regions of α -helix will often be recognizable. In many cases, the boundary of the molecular envelope will be visible, as may the symmetry relations of the subunits within an oligomeric complex.

At medium resolution (~ 3 Å), it is usually possible to follow most, if not all, the polypeptide chain, and, except in the more disordered regions on the protein surface, the electron density of sidechains is visible. At higher resolution (~ 2 Å), the peptide and side chain conformations are very much clearer. For example, the backbone carbonyl oxygens become visible, the depression in the centre of aromatic rings is evident, and the two sulphur atoms within a disulphide bond are generally resolvable.

Once a map of suitable quality and resolution is obtained, the task of model building can begin. The polypeptide chain must be traced, normally by inspection of the contoured electron density as plotted on a series of suitably spaced transparencies, and coordinates, of at least the α -carbon atoms, need to be measured. Modern computer graphics can then be employed to fit an atomic model to the observed density. This process can be further facilitated by the availability of the amino acid sequence once that sequence is aligned to the chain trace. At this stage, one uses amino acid residues with standard bond lengths and angles, and only the dihedral angles are altered to achieve a reasonable fit. It has, however, become clear that detailed understanding of specific protein

interactions requires a more accurate knowledge of the atomic coordinates. Methods have therefore been developed to refine the protein model against the experimental observables.

There are two principal techniques for the refinement of atomic coordinates and thermal parameters. The first is real space refinement, such as that proposed by Diamond (1971, 1974), which seeks to find the best fit of a model to the observed electron density by minimizing $(\rho_{\text{obs}} - \rho_{\text{calc}})^2$, where ρ_{calc} is the electron density calculated from the atomic model. The second (and generally favoured) method is the reciprocal space approach (e.g. Konnert and Hendrickson (1980)), which minimizes the function $\sum_{hkl} W (F_{\text{obs}} - F_{\text{calc}})^2$. The advent of fast Fourier transform algorithms (Agarwal (1978)) has significantly reduced the relatively large computing time necessary for the structure factor calculations involved in this latter method.

In the early stages of refinement, it is advisable to treat the MIR phases, in addition to the structure factor amplitudes, as observables. In the later stages, however, phases calculated from the atomic positions may be used in combination with, or ultimately, in place of, the MIR phases. In addition, since native protein data often extends to a resolution higher than that of the MIR phases, the resolution of the electron density map may be extended by using calculated phases for the higher order terms.

The problem of an insufficiently high ratio of observed data to refined parameters has led to the introduction of methods to ensure convergence in the refinement. These fall into two categories, either energy restraints are imposed on the structure, e.g. Jack and Levitt (1978), or bond angles and lengths are constrained to an idealized amino acid geometry (Diamond (1971), Konnert and Hendrickson (1980)).

All the methods employed in structure refinement require some manual intervention at various stages, usually in the form of difference Fourier syntheses (Deisenhofer and Steigeman (1975), Sielecki et.al. (1979)), in order that errors in the model, such as the mis-assignment of amino acids or the incorrect positioning of side chains, may be corrected. In addition, ordered solvent molecules which play an important role in the structure can be identified.

1.3 Characteristics and applications of synchrotron

X-radiation

The emission of electromagnetic radiation from accelerated charged particles has been the subject of theoretical studies for many years and has been reviewed by Jackson (1975). The physical properties of synchrotron radiation (SR) are well understood (and are calculable from the basic machine parameters): these have been reviewed by Winick and Bienenstock (1978) and by Materlik

(1982). The Daresbury Laboratory Synchrotron Radiation Source (SRS) will serve as an example of the configuration of a totally dedicated source (figure 1.10).

At the SRS, electrons are accelerated in a linear accelerator to an energy of 10-15 MeV and are then fed into a booster synchrotron where they are accelerated to 600 MeV before injection into the main storage ring operating at energies up to 2 GeV. The radiated energy loss is replenished in RF cavities operating at 500 MHz, but the electron current lifetime is limited by collisions with the residual gas in the evacuated storage ring. The electrons are held in a closed orbit by 16 dipole bending magnets, each of radius $R = 5.56 \text{ m}$ and magnetic field $B = 1.2 \text{ T}$. From each point in the curved orbit, the relativistic electrons emit SR into a very small angular cone in the forward direction so that as the electrons traverse the full bending magnet arc, the cone of radiation sweeps out a fan in the horizontal orbital plane. A number of experiments can then be positioned to accept different portions of the radiation fan, for example the protein crystallography station 7.2 'sees' the central 4 mrad of light from bending magnet 7.

In addition, multipole magnets, known as wigglers and undulators, inserted into straight sections between the bending magnets, can induce complex electron orbits to give radiation with a modified spectral profile and increased brightness compared to that from a standard bending magnet.

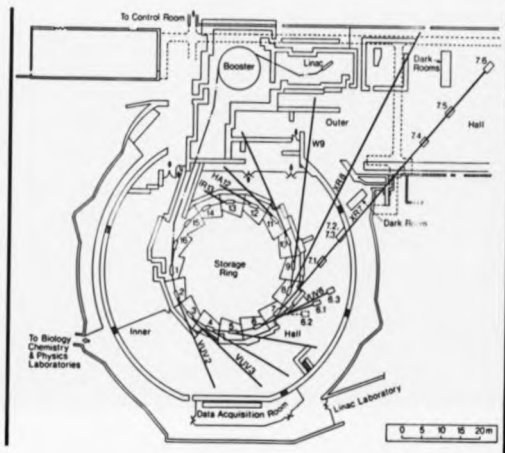


Figure 1.10 Schematic outline of the Synchrotron Radiation Source (SRS) at the Daresbury Laboratory.

1.3.1 Properties of SR

The total optical power radiated from a synchrotron with a circulating electron current of I amps, is given approximately by $88.5 E^4 I / R$. For the SRS operating at 2 GeV and 370 mA, this corresponds to 95 kW for the full 2π radians or 15 W per milliradian. The spectral distribution can be characterised by the critical wavelength, λ_c , either side of which, half the total power is radiated. Figure 1.11 shows the spectral distribution for the SRS in terms of the number of emitted photons/second/horizontal mrad/mA stored current/0.1% bandwidth. This curve is of similar shape for all SR sources and is referred to as the 'universal curve'.

The source brightness is degraded by the finite size and angular spread of the electron beam. The source size is dependant on the beam focusing magnet configuration, and, at the SRS, is calculated to be $13.7 \times 0.43 \text{ mm}^2$ - FWHM values in the horizontal and vertical respectively. The electron beam angular emittance is, in practice, considerably less than the opening half angle of the instantaneous radiation emitted by each electron; this is given by mc^2/E (where mc^2 is the rest energy of an electron) and for the SRS at 2 GeV has a value of 0.28 mrad.

In the plane of the electron orbit, the emitted radiation is 100% linearly polarised in that plane. However, deviations of the electron orbit from the mean plane give rise to a radiated component with a polarisation

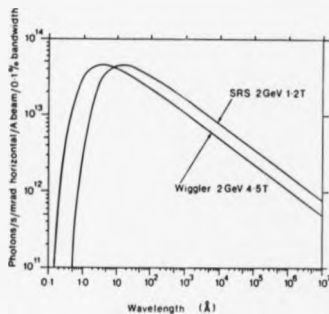


Figure 1.11 Vertically integrated spectral curves for a normal bending magnet (1.2 T) and for a 4.5 T wiggler magnet, with a 2 GeV 1 A beam in the storage ring.

perpendicular to the orbit plane. The intensity of this component is relatively small since the vertical deviations of the electrons are small.

The electrons in a synchrotron are not circulating in a continuous stream but are maintained in discrete bunches. The bunch width is determined by the RF of the linear accelerator and the minimum bunch separation is determined by the RF of the storage ring. At the SRS, in 'multibunch' mode, 160 bunches circulate with a bunch width of 0.2 nsec and a separation of 2 nsec. The machine is devoted, for a significant fraction of the time, to 'single bunch' operation with a 0.2 nsec pulse occurring every 320 nsec. Such a time structure is utilised in, for example, time resolved spectroscopic studies.

The superconducting wiggler magnet at the SRS consists of three alternating poles of field strength 2.5, 5.0, and 2.5 T respectively, thereby producing a horizontal oscillation of the electron beam path. The principle effect is to shift the spectral distribution to shorter wavelengths, significantly enhancing the photon flux available at short ($< 1 \text{ \AA}$) X-ray wavelengths. In addition, the smaller radius of magnet curvature results in an increase in radiated power: approximately 44 W/mrad at a machine current of 270 mA. However a wiggler affects the general performance of a storage ring in that injection current capacity is reduced and the source emittance is degraded to some extent.

1.3.2 Applications of SR in protein crystallography

Synchrotron radiation, with high intensity available at wavelengths ranging from infra-red to X-ray, has found wide application in surface science, material science, molecular and atomic spectroscopy, as well as structural biology. At the SRS, experimental stations have been built to cater for studies in the following areas: infra-red and time resolved spectroscopy; fluorescence lifetimes; photoionisation, photoelectron and angle dispersed photoelectron spectroscopy; extended X-ray absorption fine structure (EXAFS); soft X-ray and surface EXAFS; powder diffraction; fibre diffraction; low angle solution scattering; X-ray topography and interferometry. New experiments, currently under development, include time resolved small angle scattering and soft X-ray microscopy. The facility for protein crystallography on beam line 7 was one of the first at the SRS and a second station, on the wiggler beam line, has recently been commissioned. Progress on all the experiments are published each year in the appendix to the Daresbury Annual Report.

The applications of SR in protein crystallography fall broadly into two classes, the first of which is the exploitation of the high intensity and collimation of the X-ray beam. The problems of large data volume and inherent weakness of reflections are greatly reduced, so much so that a data collection which would span months in the home laboratory may be completed in a matter of

hours using SR. Wilson et.al. (1983) present two photographs from phosphorylase b (97000 daltons/asymmetric unit); the first was exposed for 13 hours on a Elliot G:8 rotating anode source; the second was taken in 6 minutes at LURE, Paris using SR and clearly shows a significant signal to noise improvement for the 2 Å resolution data towards the edge of the film. The lack of diffracted intensity at this resolution on the conventional source is due to the severe radiation damage suffered by the crystal, yet, at LURE, a crystal yielded five useable photographs, so enabling partial reflections from contiguous photographs to be summed together. Thus, fewer crystals are needed per dataset and more information is retrieved from each photograph.

The increase in effective crystal lifetime has consistently been reported as a benefit of SR (and the data collection described in §4 provides another example). It has become clear that radiation damage is, to some extent, dose-rate dependent in that a sample can withstand a larger total dose if the dose rate is higher. The small fraction of the X-ray photons, which are absorbed by protein crystals, produce the free radicals which are considered to be the most potent agents for damage. The dose-rate dependence may well be related to the diffusion rate of free radicals, although it is clear that the mechanisms of radiation damage are not well understood.

The study of very large unit cells presents particular

difficulties, not only are samples highly radiation sensitive and weakly diffracting, but also the reflections are very closely spaced. The high natural collimation of SR in the vertical, plus the facility to slit the horizontal convergence down to less than 1 mrad, keeps the spot size small and allows order to order resolution to be preserved. For example, rhinovirus crystals (Rossman and Erikson (1983)) are orthorhombic with cell dimensions of $320 \times 351 \times 377 \text{ \AA}$. At the SRS, it was possible to record one 3 \AA resolution photograph per crystal with an exposure time of 20 minutes. The orientation of the crystal was obtained from still photographs taken after the data photograph in order to make the best use of the fresh crystal.

The high intensity of SR also encourages the collection of data from smaller crystals than is tractable with a conventional source. Experiments with microcrystals have, to date, concentrated solely on relatively small molecules. For example, Eisenberger et.al. (1984) have measured reflection profiles from an 800 \mu m^3 zeolite crystal using SR, and Andrews et.al. (1985) have recorded complete data sets at the SRS from di- and tri- saccharide crystals measuring typically $20 \times 30 \times 250 \text{ \mu m}^3$. The major difficulties encountered are those of high background scatter and of small spots on the film, although the latter would not be a problem with an electronic area detector. Helliwell (1984) suggests that a $(50 \text{ \mu m})^3$ protein

crystal with a $(100 \text{ \AA})^3$ unit cell should be suitable for data collection on the wiggler station at the SRS but, as yet, no such experiment has been attempted.

Returning to more typical protein crystal volumes, it is possible, in certain favourable cases, to collect a complete data set quickly enough to enable study of intermediate states of structural interactions (e.g. enzyme-substrate reactions). Suitable control of the turnover rate of the reaction products, perhaps at low temperature (e.g. Alb r et.al (1976)), coupled with rapid SR data collection could lead to electron density maps of 'frozen' intermediate states. Phosphorylase b is a candidate for such a technique, a complete 3 \AA resolution data set has been collected at the SRS from a single crystal within 35 minutes elapsed time, and further studies are in progress (Medju et.al. (1987)).

An alternative approach to time resolved crystallography, which may prove to be the more useful biologically, is to use the full 'white' SR beam to record Laue diffraction patterns. This method, in its most simple geometry, uses a stationary detector. Individual reflections are fully stimulated by integration over wavelength rather than rocking width. With a broad continuous range of wavelengths in the incident beam, a large region of reciprocal space can be surveyed in a single exposure; indeed, for high symmetry hexagonal or tetragonal crystal systems, a large percentage of the unique reflections to a resolution

of 2.5 or 3.0 Å can be recorded from just one or two suitable orientations of the crystal. Using the vertically focused polychromatic radiation (wavelengths 0.8 - 3.0 Å) on the wiggler beam line at the SRS, Laue diffraction patterns from a number of different protein crystals have been recorded with exposure of, in some cases, $\frac{1}{2}$ of a second. (see figure 1.12 for a typical example). Similar experiments have also been performed at the Cornell University CHESS synchrotron using a more limited spectral range with, typically, $\delta\lambda/\lambda_{\text{mean}} = 0.2$, (Moffat et.al. (1984)).

There are various problems to be overcome in using this approach, the most limiting being, that only the most robust protein crystal will survive more than one or two exposures. Most of the difficulties encountered in processing the raw patterns, high density of spots (up to 20000 per exposure), higher order overlaps, spectral normalisation, appear surmountable and a suitable software package is under development (Helliwell et.al. (1985)). Using the Laue method, in conjunction with a fully focused white beam from a high brightness synchrotron source, offers the potential to investigate protein interactions structurally on a sub-millisecond timescale.

The second principle application of SR in protein crystallography is the utilisation of wavelength tunability in solving the phase problem. The principles of anomalous scattering, described above (§1.2.1), found regular use

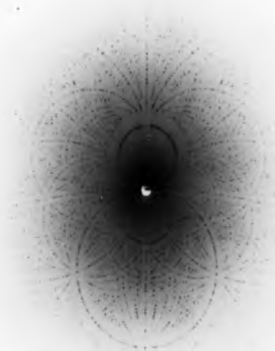


Figure 1.12 Laue pattern recorded on the SRS wiggler from a crystal of pea lectin (space group $P2_12_12_1$, $a=50.7$, $b=61.2$, $c=136.7$ Å). The exposure time with the SRS operating at 1.8 GeV, 20 mA (in single bunch mode) was 45 seconds (courtesy of Dr. J.R. Hellivell).

in phase determination long before the advent of SR. However, experiments had been limited to the discrete wavelengths available from conventional X-ray sources, whereas with SR the wavelength can be selected to probe any absorption edge of interest and optimize the anomalous signal in the diffraction pattern. Figure 1.13 shows the wavelengths of the K and L absorption edges in the relevant range of atomic numbers, and figure 1.7 showed idealised plots of the real and imaginary components of the anomalous scattering in the region of an absorption edge. The values quoted for f' and f'' in the International Tables of Crystallography, and those calculable by the method of Cromer (1983), are for pure elements. However, for bound atoms of various valence states, it is found that the exact wavelength of the absorption edge may be significantly shifted and that the local chemical environment of the absorbing atom introduces considerable 'structure' into the linear absorption coefficient. For example, the L_{III} spectrum of $K_2Pt(CN)_4$, a commonly used heavy atom compound in protein crystallography, has been measured by Hellwell, Phillips and Siddons (unpublished work) (figure 1.14). Comparison of the calculated f'' curve and the measured absorption curve enables values of f'' to be estimated for the sample under examination. In this example, the spectrum should be a good approximation to that of the $Pt(CN)_4^{2-}$ bound to the protein since the square planar

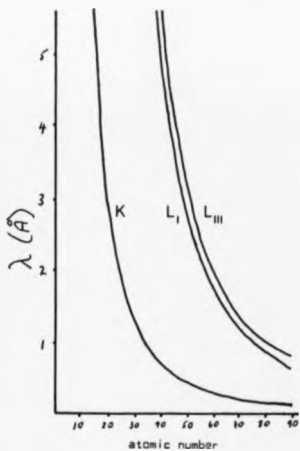


Figure 1.13 The wavelengths of K and L absorption edges, plotted as a function of atomic number.

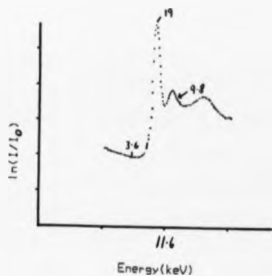


Figure 1.14 The L_{2M} absorption edge of Pt recorded from a $K_2Pt(CN)_4$ sample on DORIS using a Si double crystal monochromator with 0.7 eV bandwidth. Estimates of f'' (in electrons) are given at specific points (from Hellivell, Phillips and Siddons (unpublished)).

configuration of the cyanide ligands is preserved. There is, however, a considerable dichroism effect for Pt in the similar structure $\text{Pt}(\text{Cl}_4)^{2-}$; these effects are important and do not show in the absorption measurements of an amorphous sample (Templeton and Templeton (1985)).

In §1.2.1, it was shown how the anomalous scattering information from a heavy atom derivative could be incorporated to resolve the phase ambiguity. As figure 1.7 showed, the maximum of f'' does not coincide with the minimum in f' and so it can be seen that it would be possible to phase reflections from measurements made at two or more suitable wavelengths from a single system containing an anomalous scatterer. Figure 1.15 demonstrates this and emphasizes the importance of selecting suitable wavelengths to obtain the maximum from the available phasing power. Figure 1.16 plots f' against f'' for praseodymium and samarium for a range of wavelengths across their respective L_{III} absorption edges, and the area of the 'loop' in these plots can be thought of as being proportional to the phasing power. Phillips (1978) and Narayan and Ramaseshan (1981) have considered schemes to decide the best choice of wavelengths and they conclude that, for a three wavelength experiment, the ones to use are those which maximize the area of the triangle defined by the three points in the f', f'' plot.

The development of the polychromatic profile method to which chapter two is devoted), represents a new

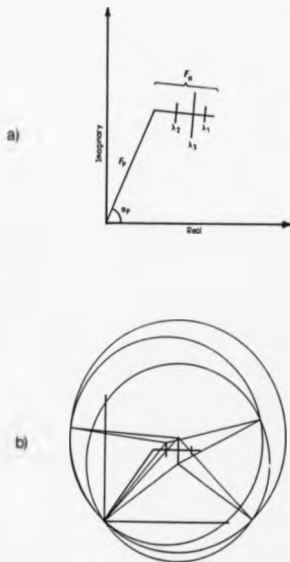


Figure 1.15 a) The contributions to a Bragg reflection (and its mirrored inverse) at three wavelengths around an absorption edge. $\lambda_1 > \lambda_2 > \lambda_3$, where λ_1 corresponds to the minimum of f'' , λ_2 to the maximum of f'' , and λ_3 to a position in the pre-edge region.

b) The centres of the phasing circles are based on the Friedel equivalent pair at the peak of the absorption, and one equivalent pair at the minimum in f' . The phase in this two wavelength, three measurement example is well resolved.



Figure 1.15 c) The centres of the phase circles are based on measurements of the same member of a Friedel pair at three different wavelengths where f'' is fairly constant. The phase remains unresolved because the centres are co-linear.

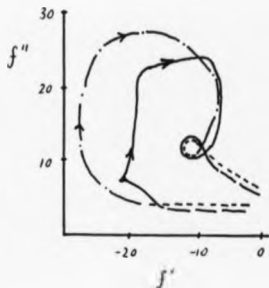


Figure 1.16 Plot of measured values of f' and f'' for Pr (---) and Sm (—) in the isomorphous crystal salts of sodium ethylene diamine tetracetate octahydrate. The direction of the arrows indicate increasing SR photon energy. The total wavelength change was 0.0088 Å for Pr and 0.0115 Å for Sm, with a $\frac{\Delta\lambda}{\lambda}$ at each point of 10^{-3} (based on tabulated data from Templeton et. al. (1980)).

approach to anomalous dispersion phasing in that it is equivalent to the sampling of f' and f'' continuously around the 'loop' of the f' , f'' plot. Therefore providing the spectral resolution at each point within the reflection profile is sufficiently high, the technique can realize all the phasing power that the phenomena of anomalous dispersion has to offer.

Such measurements place rigorous demands on the experimental instrument. The wavelength needs to be accurately controlled in order that the desired 'operating point' is obtained, and the wavelength must be stable. In addition, the wavelength spread, $\delta\lambda/\lambda$, in a monochromatic beam must be kept to a minimum to avoid a smearing of the centres of the phase circles and an effective reduction of f' and f'' . The need to scan the absorption edge in a suitable compound, modelling that in the crystal, is also clearly identified. A preferable approach would be to use a sample crystal to optimize f'' and either monitor the diffraction pattern itself using an electronic area detector, or alternatively, monitor the absorption by fluorescence measurement (Cant et.al. (1983)).

The topic of area detectors (dealt with in the following section) is central to the full exploitation of optimized anomalous scattering; anomalous differences in the diffraction pattern are generally small and the signal to noise ratio obtained using film is marginal. The full power of the method is not likely to be realized

until electronic area detectors are operating routinely at SR sources, although a successful three wavelength experiment has been recently reported by Kohn et.al. (1985). They determined phases to 2.3 \AA resolution for a terbium parvalbumin using only anomalous scattering at wavelengths around the Tb L_{III} absorption edge. The resulting structure shows good agreement with that previously determined by molecular replacement methods.

1.3.3 Experimental considerations

This section covers the X-ray optics and instrumentation required for SR protein crystallography and, in particular, describes the experimental facilities at the SRS. The design and performance of the first protein crystallography station at the SRS (designated 7.2) has been reported by Helliwell et.al. (1982a), while the mechanical arrangement and stepper motor control software has been fully described by Carr (1984). A description of the relatively new station on the wiggler beamline (designated 9.6) has been given by Helliwell et.al. (1986).

To make full use of the available flux, the SR beam must be focused onto the sample, and, at the SRS, the source size to be imaged is $13.7 \times 0.43 \text{ mm}^2$ (§1.3.1). In the vertical direction, 1:1 focusing is adequate and this is achieved by a 582 mm long, by 80 mm wide, by 30 mm deep, polished fused quartz mirror

coated under UHV with platinum, and situated 11 m from the source. By applying a couple at each end, the mirror is bent into a cylindrical curvature with a radius of approximately 1 km, to provide the necessary focusing. The angle of the mirror with respect to the beam is set such that wavelengths below 1 \AA are rejected ($1/2 \text{ \AA}$ on the wiggler station).

The horizontal characteristics of the beam are left unaltered by the mirror, and horizontal focusing, requiring 10:1 demagnification, to better match the focal size to the sample size, is provided by bending the monochromator crystal. Perfect crystals of silicon or germanium have peak reflectivities approaching 100% in the wavelength range of interest ($0.5 - 3.0 \text{ \AA}$), and can be grown to sizes large enough that the full 4 mrad divergence can be accepted. Monochromatisation is achieved by reflecting the radiation from the (111) crystal plane so that Bragg's law is only satisfied by one particular wavelength. By rotating the crystal and so altering the angle of incidence the wavelength is tunable over a range limited only by the physical restrictions on the 2θ angle ($0 - 70^\circ$ on station 7.2). The triangular shape of the crystal ensures a cylindrical curvature when a couple (provided by an ascentric cam) is applied to the tip. Horizontal demagnification is provided by cutting the crystal surface at an angle, α , to the reflecting planes. All aspects of this optical

system are dealt with fully in §2.1. A schematic of the optics at the SRS is shown in figure 1.17.

An alternative configuration of the X-ray optics would employ a double crystal monochromator in which the emergent beam is parallel to, and only slightly displaced from, the incident beam for a wide range of reflected wavelengths (figure 1.18). The two crystals can be separate and individually controlled, or coupled together in a monolithic channel cut device. Such a system, in a vertically dispersing mode, would provide rapid tunability, and, with suitable vertical entrance slits, high spectral resolution. Horizontal focusing can then be achieved by using a doubly focusing mirror, although with a large horizontal source size, as at the SRS, the intensity at the sample would be significantly less than with the singly focusing mirror plus the asymmetric cut, singly bent triangular monochromator. The double crystal arrangement is ideal for optimised anomalous scattering experiments, and it is planned that this option be installed on the wiggler station at the SRS.

The X-ray camera is securely mounted on top of an alignment carriage which has 4 degrees of freedom, two rotational and two translational about axes perpendicular to the X-ray beam, provided by stepper motors which are capable of 5 μm movements per step. The carriage is mounted via recirculating roller bearings on a 4 m experimental arm which extends horizontally from below

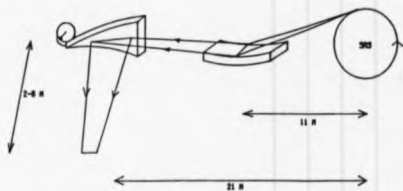


Figure 1.17 Schematic of the X-ray optics for protein crystallography at the SRS.

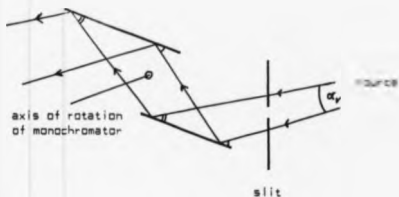


Figure 1.18 Double crystal monochromator in the vertically dispersing mode. The range of wavelengths reflected is determined by the SR vertical divergence angle, or the angle subtended at the source by an entrance slit, whichever is the smaller.

the monochromator vessel. The carriage can therefore be moved along the arm to match the focusing distance of the monochromator. The experimental bench is supported by three thrust bearings; one is directly below the monochromator mount and provides the pivot point about which the arm can rotate, and the other two support the arm, via bushes, on horizontal cross track rails which are permanently attached to the floor. This arrangement results in a large stable bench which can be easily moved by hand through a 20 range of 70°, limited only by the extent of the cross tracks. In addition to the alignment carriage, the experimental bench also supports an evacuated beam pipe which reduces the air path between the monochromator vessel and the camera to a minimum. Finally the entire instrument is enclosed in a radiation safety hutch designed to ensure no access during an X-ray exposure. The various stepper motors used for the adjustment of carriage, monochromator, mirror and slits are all controlled from outside the hutch by a dedicated LSI 11/23 computer via a CAMAC based motor address system. A photograph of the complete instrument on station 7.2 is shown in figure 1.19.

In anomalous dispersion experiments, where the signal of interest is relatively weak, the accuracy of the intensity measurements becomes doubly important. Photographic film has long been the only area sensitive



Figure 1.19 Photograph of the instrument on station 7.2 at the SRS.

detector available, and, in many respects, it remains the most flexible and convenient detector to use. Its principle advantage is excellent spatial resolution, limited only by the size of a single grain (a few μm), and, thus, it meets the needs of the trend to study larger and larger unit cells. However, film does not allow on-line control of the experiment and, most importantly, it is relatively inefficient. To produce reasonable blackening on a film (say one optical density requires 10^6 photons/ mm^2 (Morimoto and Uyeda (1963)) or about 10^5 photons in a typical 0.3 mm diameter spot, which would result in reflection agreement statistics of, at best, 5%. In a photon counting system, 10^5 photons would give a statistical counting error of 0.3%, or, alternatively, 5% accuracy would be achieved with less than 1000 photons. It is for this reason that, providing the spatial resolution is not limiting, electronic area detectors can offer real advantages in the accurate measurement of intensities.

Multiwire proportional chambers (MWPC) have been successfully used with conventional sources for structure determination (Matthews et.al. (1977), Xuong et.al. (1978) Xuong et.al. (1985)) and are under development at most SR sources, including the SRS (Helliwell et.al. 1982b). Such detectors have a resolution of 1-2 mm, suffer from parallax depending on the chamber depth, and have a limited maximum counting rate of $10^5 - 10^6$ counts

per second. However, for weakly diffracting samples or for optimized anomalous dispersion experiments, where flux is sacrificed for spectral purity, MWPCs are highly attractive for use with SR.

TV image intensifier systems are also being developed for protein crystallography (Arndt and Gilmore (1979), Arndt and Thomas (1982), Kalata (1982)). A FAST TV diffractometer system supplied by Enraf Nonius will be commissioned at the SRS on station 9.6 during 1985. With a high count rate (10^8 per sec) and reasonable resolution (0.5 nm) these devices combine some of the advantages of film with the single photon counting statistics of electronic detectors. With the advent of these modern systems it is likely that, for experiments where accuracy of intensity measurements is important, film will be rendered obsolete.

1.4 Drug design and crystallography

Prior to the work of Paul Ehrlich, who in 1910 developed Salvarsan as a treatment for syphilis (Kunin (1970)), drugs were discovered by testing extracts of natural products for useful properties. Since that time, drug development has relied upon the empirical testing of a large number of chemicals for a desired biological response in model systems, followed by systematic modification of a lead compound to optimize its properties. In view of the now apparent complexity of biological systems, it is not surprising that this 'trial and error' approach has proved highly inefficient and expensive. In addition, many drugs that have found

their way into widespread use have either a very low therapeutic activity, or, as in cancer chemotherapy, cause serious side effects.

In recent years, the concept of a rational approach to drug design has been developed around the advent of structure - activity - relationships (SAR), which allow biological properties to be quantitatively correlated with the physiochemical or structural properties of chemical components. Austel (1984) has outlined the principles behind a systematic and practical strategy in the search for new drugs, starting with the determination of a structural field around a lead compound and proceeding to optimization using, initially, qualitative SAR and, ultimately, quantitative SAR, in which a biological property is expressed in terms of a mathematical function of physiochemical and structural descriptors. In this framework, the experiments of the medicinal chemist should be designed to elicit a maximum of SAR information.

Although conformation is not the sole determinant of a drug interaction (lipophilicity, basicity and charge distribution are also significant influences), it is clearly of major importance. Many drugs act on target receptors in a stereoselective fashion to elicit pharmacological effects. It is often found that one isomer of a drug is more active than another in a certain test situation, while the activity values are reversed in different tests (e.g. the adrenoceptor blocking potency of labetalol (Brittain et.al. (1982))).

Alternatively, all isomers may be nearly equipotent in their action on a given receptor, yet one can show toxic side effects, e.g. Ketamine (White et.al. (1980)) and thalidomide (Blaschke et.al. (1979)). In addition, it is highly probable that a single conformation only of a flexible drug molecule is bound in the drug-receptor complex. Increasing emphasis has therefore been laid on the elucidation and analysis of configurational and conformational characteristics of drugs and their receptors by x-ray crystallography, nuclear magnetic resonance (NMR), and quantum chemical calculations. It must always be remembered that an x-ray crystal structure is merely an averaged static model of a possibly flexible molecule influenced by crystal packing forces. However, crystallographic data is generally in very good agreement with spectroscopic data. Also, evidence is beginning to emerge from empirical force field calculations, showing good agreement between the conformation observed crystallographically and that preferred energetically (e.g. the comparison of steroid structures by Duax et.al. (1981)).

There are many examples where the crystal structures of families of related drugs have contributed to the understanding of drug-receptor interactions. Often in these cases, the receptor is unidentified, let alone isolatable. Camerman and Camerman (1984) have studied crystallographically a wide range of chemically diverse compounds which have a single common pharmacological action, namely anti-convulsant activity, thereby

establishing the stereochemical features common to the three dimensional conformation of these compounds, and enabling them to suggest guidelines for the design of new potential antiepileptic drugs.

Most of the drugs used to combat human cancers are ultimately active at the DNA level. They either modify DNA metabolism or interact directly with the DNA molecule, so interfering with the genetic apparatus of cells, particularly rapidly proliferating ones. The favoured model, suggested by Lerman (1961), for the latter class of mechanism, is one of intercalation. Crystal structures of model complexes (notably drug-duplex dinucleoside monophosphate) have revealed some of the subtleties of the complex range of intercalation phenomena and have contributed significantly to the knowledge of SAR. This work has recently been reviewed by Neidle and Berman (1983), and by Neidle (1984).

In cases where the protein receptor is identified and available for study in the crystalline state (or has a known sequence which is homologous with related proteins that do readily crystallize), protein crystallography offers the opportunity to study protein-inhibitor complexes at atomic or near-atomic resolution. Modern interactive computer graphics techniques, capable of displaying complete proteins as line drawings or as 'dot' or 'net' surfaces (e.g. Tickle (1982), Pearl and Honegger (1983)), can then be used to model the complex, generate complementary surfaces and examine docking of molecules.

One of the most widely studied drug receptors is dihydrofolate reductase (DHFR). The anti-cancer properties of methotrexate (MTX), the anti-bacterial properties of trimethoprim, and the anti-malarial properties of pyrimethamine are all related to inhibition of DHFR, thus emphasizing the important metabolic role and wide distribution of the enzyme (see Hitchings and Smith (1980) and Beddell (1984) for general reviews). DHFR (with reduced nicotinamide adenine dinucleotide phosphate (NADPH) as cofactor) reduces dihydrofolate (DHF) to tetrahydrofolate, which is then converted to N⁵, N¹⁰ - methylenetetrahydrofolate, a necessary cofactor in the synthesis of thymidylate from deoxyuridylate by thymidylate synthetase (itself an important target for chemotherapy). Thymidylate is essential for DNA synthesis and so inhibition of its production results in cell death. Since the first X-ray crystal structure, that of the E.coli DHFR-MTX complex by Matthews et.al. (1977), a variety of DHFRs from bacterial and mammalian sources have been studied.

There is strong evidence from the X-ray structures, supported by ¹³C NMR studies (Cocco et.al. (1983)), that N1 of MTX is protonated when bound to that enzyme, and that a charge interaction exists between an aspartate (bacterial) or a glutamate (vertebrate) and the protonated N1. The refined crystal structures at 1.7 Å resolution (Bolin et.al. (1982)) of the E.coli and L. casei DHFR-MTX complexes, show that since MTX binds with the pteridine ring in a flipped orientation as

compared to the natural substrate, DHF, an additional hydrogen bond forms at the 4-amino which cannot form for DHF. Using site directed mutagenesis, a mutant E. coli DHFR has been expressed in which the key aspartate was replaced by arginine (Villafranca et.al. (1983)). The mutant enzyme shows no protonation of MTX in the binary complex, but only a 100 fold lower binding constant for MTX than the wild type enzyme, thus indicating that the charge interaction is not wholly responsible for the $10^4 - 10^5$ higher affinity that DHFR shows for MTX over DHF. These, and other results, are guiding the efforts of the various pharmaceutical laboratories concerned with the development of new, more effective inhibitors of DHFR.

The DHFR example shows the role protein crystallography can play in identifying the important structural interactions between drug and receptor. When coupled with other experimental or theoretical techniques (e.g. charge densities given by quantum mechanical calculations can be used to generate electrostatic contour maps), crystal structures provide a sound basis for the rational design of new drugs which should supercede the guesswork of yesterday.

Purine nucleoside phosphorylase (PNP) is one of the first proteins to be selected for crystallographic structure determination primarily on the basis of the potential therapeutic value of an improved inhibitor of the enzyme,

following the failure of standard drug design methods to produce a suitable compound. Interest in the development of effective PNP inhibitors has been generated by knowledge of the relationship between PNP deficiency and certain immunological diseases, and of the role of PNP in the metabolism of purine nucleosides. The results of the extensive research into the function and characteristics of PNP are reviewed in §3 of this thesis. Since the enzyme binds various substrates and substrate analogues, it should be possible to use these complexes to characterize the active site and identify the nature and spatial arrangement of the functional groups which are important for the enzymatic activity. In addition, the crystal structure may also suggest entirely new regions on the enzyme that could be potential targets for compounds which would affect the enzyme by allosteric mechanisms.

CHAPTER TWO

THE SR POLYCHROMATIC PROFILE METHOD IN CRYSTALLOGRAPHY

Optical systems based on a singly bent triangular perfect-crystal monochromator have been constructed at LURE (Lemmonier et.al. (1978); Kahn et.al. (1982)), DORIS (Hendrix et.al. (1979)), and the SRS (Helliwell et.al. (1982a)) to provide intense X-ray beams suitable for single crystal diffraction experiments. The first half of this chapter describes the properties of such a system relevant to the diffraction process and the single crystal oscillation method. In doing so the possibility of a polychromatic method in which an energy profile is produced along the axis of each diffraction spot in a screenless oscillation photograph, is developed. The second half of the chapter describes experiments conducted at the SRS which demonstrate firstly, that optimized anomalous dispersion experiments requiring high spectral purity can be catered for; and secondly, that the polychromatic profile method is feasible.

All the material covered in this chapter has been previously published. Greenhough and Helliwell (1982a, 1982b) give detailed descriptions of the diffraction geometry for both conventional and synchrotron sources. Greenhough, Helliwell and Rule (1983) derive the general

diffraction spot size, shape and energy profile, and illustrate these with particular examples from the polychromatic experiment described in §2.2.2. Arndt, Howard, Greenhough, Helliwell, Rule and Thompson (1982) describe the necessary experimental conditions and the more general results of that experiment.

2.1 Theory

2.1.1 The monochromator

The use of a bent triangular monochromator crystal was briefly described in §1.3.3. Provided the curvature can be accurately controlled, the bending radius can be set such that the diverging 'white' rays from a point source all make the same angle of incidence along the curved surface and hence, the same photon energy is reflected from all illuminated points along the monochromator length. This is the 'achromatic condition' of the monochromator and is commonly referred to as the 'Guinier position'. In the case where the reflecting Bragg planes are parallel to the crystal surface, the achromatic condition corresponds to a focusing distance p' equal to the source to monochromator distance p and to a horizontal image size h' equal to the horizontal source size h . This arrangement would generally be unsuitable for current sources of SR because p is usually greater than 10 m and source sizes are considerably larger than typical protein crystals.

By cutting the monochromator crystal at an angle, α , to the reflecting planes the reflected beam is compressed in width compared to the incident beam (figure 2.1).

The compression ratio, A , is given by

$$A = \frac{\sin(\theta - \alpha)}{\sin(\theta + \alpha)},$$

where θ is the Bragg angle of reflection and, for α positive and $\theta > \alpha$, $A < 1$ (Lemmonier et al. (1978)). The image distance p' at the Guinier position is Ap and the focal width is Ah . A Ge(111) crystal with an oblique cut of 10.5° set to reflect 1.5 \AA radiation, has a compression ratio of $\sim 1/7$.

If the oblique cut is zero then the monochromator acceptance angle, ω_{acc} , and the angular rocking width of the reflected beam, ω_{diff} , are both equal to ω_{sym} , the monochromator rocking width. For a positive ω_{diff} is increased by $A^{-1/2}$ and ω_{acc} is reduced by $A^{1/2}$, thereby reducing the total flux by $A^{1/2}$. However, the focal width is reduced by A and so the final horizontal flux density is improved by $A^{-1/2}$. In addition, the geometrical width of the monochromator, seen by the incident beam, is increased from $L \sin \theta$ to $L \sin (\theta + \alpha)$ where L is the length of the monochromator. This results in an increased flux if the width of the available beam is $> L \sin \theta$.

The spread of wavelengths $(\delta \lambda / \lambda)_{tot}$ in the reflected beam is determined by several factors including the crystal rocking width and any variation in the angle of incidence along the length of the monochromator. Figure 2.2 shows

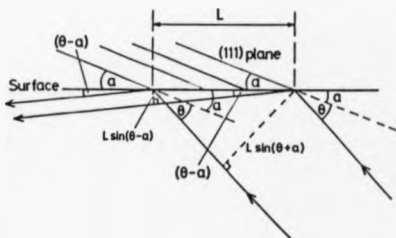


Figure 2.1 Reflection from an asymmetrically cut crystal monochromator. The width of the reflected beam is compressed relative to that of the incident beam.

how the finite horizontal source size produces a photon energy gradient across the focus. At the Guinier position, all rays from a point in the source have equal angles of incidence, but rays from different points will be incident at slightly different angles and so different wavelengths will be reflected. Hence there is a correlation between wavelength and position across the focused image. By placing a narrow slit ($\ll \lambda/p$) at the focus a fine spectral resolution beam can be selected (and is equivalent to placing a slit at the source itself). This is demonstrated by the Dy_2O_3 powder experiment described in §2.2.1.

Deviation from the Guinier position (i.e. the monochromator is 'underbent' or 'overbent') also leads to an increased spread in wavelengths reflected. Figure 2.3 demonstrates the effect in the absence of other considerations. In this case, the wavelength is correlated with direction rather than position in the horizontal line focus. Helliwell et.al. (1982) placed an Fe foil directly behind a narrow slit at the focus and, with radiation from an overbent monochromator, recorded photographically the Fe K absorption profile along its length. It is this principle that is exploited in the polychromatic diffraction experiments described in §2.2.2.

2.1.2 Spectral resolution

The contribution to the total wavelength spread due to the 'curvature component' described above, is

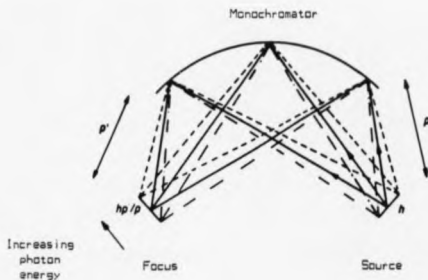


Figure 2.2 All rays from a point in the source are incident at the same angle to the monochromator when at the Guinier position. Rays from different points in the source are incident at different angles, and different wavelengths are reflected.

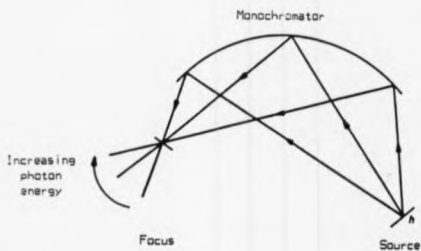


Figure 2.3 Away from the Guinier position, rays from a point in the source to different points on the monochromator are incident at different angles. At each point in the focus the direction of the rays are correlated with wavelength.

$$\frac{L}{2} \left| \frac{\sin(\theta-a)}{p'} - \frac{\sin(\theta+a)}{p} \right| \cot \theta$$

(Lemmonier et.al. (1978)). At the Guinier position $\sin(\theta-a)/p' = \sin(\theta+a)/p$, and the expression reduces to zero. The source size contribution to the spectral spread is $\delta\theta \cot \theta$ where $\delta\theta = h/p$. These geometrical factors are then convoluted with the crystal rocking width to obtain the total wavelength spread across the focus.

The relevant contribution here is $\omega_{acc} \cot \theta$; towards the centre of the focus, however, $\omega_{diff} \cot \theta$ ($> \omega_{acc} \cot \theta$) defines the energy resolution along each ray (Kohra et.al. (1978)).

Thus we have,

$$\left(\frac{\delta\lambda}{\lambda} \right)_{TOT} = \left\{ \left(\frac{h}{p} \right)^2 \left(\frac{\sin \theta - a}{p'} - \frac{\sin(\theta+a)}{p} \right)^2 + \omega_{acc}^2 \right\}^{1/2} \cot \theta.$$

If, however, the sample (or slit) is less than the full focal width then the source size contribution $(h/p) \cot \theta$ reduces to $(x_H/p') \cot \theta$ where x_H is the horizontal sample size.

In addition to the total spectral spread across the focus there are limitations to the spectral resolution at a point within the focus. Each incoming ray has a spectral spread $(\delta E/E) = \omega_{diff} \cot \theta$ (see above) due to the monochromator rocking width. The resolution would therefore be improved by using a symmetric cut crystal (i.e. $a=0$). At the SRS for 1.743 Å (Fe K edge) radiation, $\delta E = 5.1$ eV for a Ge(111) crystal ($a = 10.44^\circ$, $\omega_{sym} = 16''$, $\delta E/E = 7.2 \times 10^{-4}$, $(\delta E/E)_{sym} = 3.2 \times 10^{-4}$) and $\delta E = 1.9$ eV for a Si(111) crystal ($a = 10.29^\circ$, $\omega_{sym} = 6''$, $\delta E/E = 2.7 \times 10^{-4}$, $(\delta E/E)_{sym} = 1.2 \times 10^{-4}$). The values quoted for a

and for w_{sym} were measured by Helliwell (1981) and Clark et.al. (1981), respectively.

The cylindrical curvature geometry does not produce a perfect point focus of a point source; there is a small horizontal aberration $L^2 \cos(\theta - \alpha) / 8R$ where R is the radius of curvature of the monochromator (Martin and Cacak (1976)). For a fully illuminated ($L=200\text{mm}$) Ge(111) crystal ($\alpha=10.44^\circ$) set at the Guinier position ($R=50\text{m}$) for 1.60811 \AA (Co K edge) radiation, the aberration is only 0.1 mm . If this is convoluted with the effect of the monochromator rocking width the effective angular width of the reflected beam is only increased from $40''$ to $40.5''$.

Associated with the focusing aberration is a depth of focus equal to $(L/2) \cos \theta$. There is also a finite 'depth of source' which contributes to the depth of focus due to the finite acceptance angle of the SR beam. For a storage ring bending magnet radius of 5.5 m and 4 mrad of beam accepted, the depth of source is 22.5 mm . For $L = 200 \text{ mm}$ and Co K edge radiation, the total depth of focus is 119 mm ; for $L = 100 \text{ mm}$ it is 60 mm . In terms of energy resolution a value in the range $1-3 \text{ eV}$ ($L = 100-200 \text{ mm}$) is obtained.

The final factor to be considered is that of an error in the value of the oblique cut leading to an error in the Guinier position. Since

$$p'_g = \frac{\sin(\theta - \alpha)}{\sin(\theta + \alpha)} p,$$

where p'_g is the value of p' at the true Guinier position, then

$$\delta p'_g = - \frac{p \sin 2\theta \cdot \delta \alpha}{\sin^2(\theta + \alpha)},$$

and this error in p'_g will contribute to $\delta\lambda/\lambda$ by the curvature component giving

$$\delta \left(\frac{\delta\lambda}{\lambda} \right) = \frac{1}{2} \left(\frac{\cos(\theta + \alpha)}{p'_g} + \frac{\cos(\theta + \alpha)}{p} \right) \delta \alpha.$$

Measurement of various crystals in use at the SRS has revealed differences of up to 0.25° between the nominal and measured values (Helliwell et.al. (1982)). The estimated error in measurement was $\pm 0.05^\circ$ for which $\delta(\delta\lambda/\lambda) = 1.2\text{eV}$ (if $L = 200 \text{ mm}$).

Adding these factors together and ignoring the horizontal focusing aberration, values are obtained for the energy resolution at a point within the focal width; at the Guinier position for Fe K_α radiation with a 10.5° asymmetric cut $\text{Si}(111)$ crystal of illuminated length 200 mm , δE summed over $\omega_{\text{diff}} \cot \theta$ (1.9eV), $\delta \alpha = 0.05^\circ$ (1.2eV), and depth of focus (3eV) becomes 6.1eV ; for $L = 100 \text{ mm}$ δE is 4.1 eV . For a $\text{Ge}(111)$ crystal the equivalent values are 9.1 eV and 7.1 eV . So $\delta E/E$ is generally less than 10^{-3} and can be reduced to below 5×10^{-4} which is perfectly suitable for scanning many absorption edges in optimized anomalous scattering experiments. Away from the Guinier position $\delta \alpha$ and the depth of focus affect only the total $\delta\lambda/\lambda$ achievable and so only $\omega_{\text{diff}} \cot \theta$ need be considered in the polychromatic profile method.

Having investigated the characteristics of the beam

focus, a diffraction model is needed for the sample crystal, describing the relevant parameters of the focused radiation from the monochromator not, in general, at the Guinier position. The sample receives rays converging at an angle β_H ($\approx L/p'$) $\sin(\theta - \alpha)$ with a range of energies correlated with direction due to the monochromator curvature. In addition, there is a slight progressive change in mean direction of incidence from one side of the focus to the other, (γ_H). These features are illustrated in figure 2.4. At the Guinier position the curvature component of spectral dispersion, $(\delta\lambda)_c$, is zero, and in the underbent case $(\delta\lambda)_c$ is negative. Figure 2.5 shows the same extreme rays transferred to intercept the origin of reciprocal space. Figure 2.6 plots incoming angle against wavelength for each fan of rays. Each incoming ray direction then has an associated wavelength spread of

$$\left(\frac{\delta\lambda}{\lambda}\right) + \left(\frac{\delta\lambda}{\lambda}\right)_{\text{foc}} = \left(\omega_{\text{diff}} + \frac{\gamma_H}{p}\right) \cot\theta,$$

and this can be described as a 'conventional-source-type' term $(\delta\lambda/\lambda)_{\text{conv}}$. For the overbent case (figure 2.6a) this spread is about a mean varying from

$$\lambda_{\text{min}} = \frac{(\delta\lambda)_c}{\beta_H} \frac{\gamma_H}{2p}, = \lambda_{\text{min}} \text{ at } +\frac{\gamma_H}{2} \quad \text{to}$$

$$\lambda_{\text{max}} = (\delta\lambda)_c + \frac{(\delta\lambda)_c}{\beta_H} \frac{\gamma_H}{2p}, = \lambda_{\text{max}} \text{ at } -\frac{\gamma_H}{2},$$

giving a spread $\delta\lambda = \frac{\gamma_H}{\beta_H} (\delta\lambda)_c$;

thus, there is a direction-correlated term

$$\left(\frac{\delta\lambda}{\lambda}\right)_{\text{corr}} = \frac{\gamma_H}{\beta_H} \left(\frac{\delta\lambda}{\lambda}\right)_c$$

$$= \frac{\gamma_H}{\beta_H} \frac{L}{2} \left(\frac{\sin(\theta - \alpha)}{p'} - \frac{\sin(\theta + \alpha)}{p} \right) \cot\theta,$$

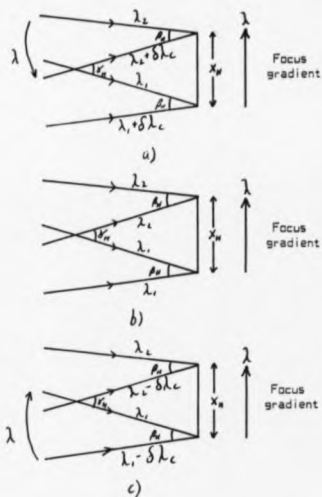


Figure 2.4 Extreme incoming rays to the sample, at the horizontal limits of the sample, at a) overbend, b) Guinier position and c) underbend. The energy gradient due to the source size effect is shown, and $\delta\lambda_c$ represents the curvature component of spectral dispersion. Use of the same symbols in each diagram does not denote equality. $\lambda_2 > \lambda_1$.

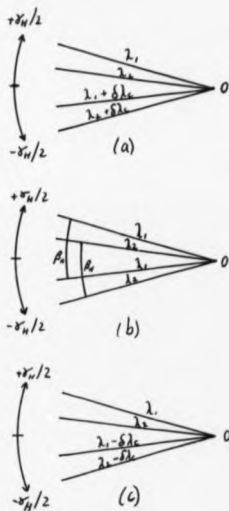
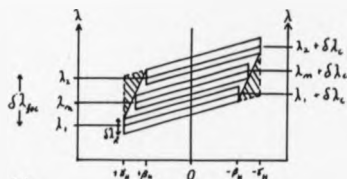
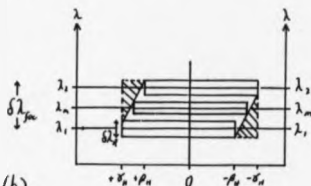


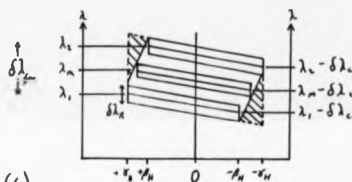
Figure 2.5 Transfer of the rays of figure 2.4 to meet the origin of reciprocal space, 0, for a) overbend, b) Guinier position and c) underbend.



(a)



(b)



(c)

Figure 2.6 Plot of wavelength versus incoming angle to the origin of reciprocal space for a) overbend, b) Guinier position and c) underbend.

$\delta\lambda_m$ is the source size component, $(x_H/p')\cot\theta$; $\delta\lambda_R$ is the rocking width component, $\omega_{H\theta}\cot\theta$; $\delta\lambda_c$ is the curvature component. $\delta\lambda_m$ and $\delta\lambda_R$ are grossly enlarged with respect to $\delta\lambda_c$.

In terms of the Ewald sphere, the direction-correlated wavelengths in the horizontal plane give rise to Ewald spheres of radius 1 along the central ray (i.e. $d^* = \lambda/d$), of radius $\lambda/(\lambda - (\gamma_H/B_H)\delta\lambda_c/2)$ along $\gamma_H/2$, and $\lambda/(\lambda + (\gamma_H/B_H)\delta\lambda_c/2)$ along $-\gamma_H/2$.

In the vertical direction there is no correlation between wavelength and direction. In the absence of a vertically focusing mirror, the angular spread is determined by the vertical beam divergence, α_V , and the source height, h_V , so that for a sample (or slit) of height x_V at the focus, the accepted divergence is $\gamma_V = (x_V + h_V)/(p + p')$. At the SRS $h_V = 0.43$ mm (calculated) FWHM and $\gamma_V = 0.03$ mrad for a 0.3 mm sample. The use of a 1:1 vertically focusing mirror reproduces α_V at each point in the focus, so that $\gamma_V = \alpha_V$, which is 0.25 mrad at the SRS.

Greenhough and Helliwell (1982b) have given tables of the parameters discussed above for various geometries available on the protein crystallography workstation at the SRS.

2.1.3 Reflecting range and prediction of partiality

The asymmetry in the diffraction model introduced by the direction-correlated spectral dispersion leads to large changes in the crystal sample reflecting range and cannot be ignored in the processing of films nor in the online prediction of reflections for electronic area detector systems being developed at SR sources.

In the case of a conventional source the symmetric spectral dispersion can simply be added to the sample mosaic spread (η) (Greenough and Helliwell (1982a)) so that the reflecting range of a reflection R , is

$$\theta_R = L(\eta d^* \cos \theta + \frac{\delta \lambda}{\lambda} d^* \sin \theta)$$

in the absence of beam cross fire. Each effect in turn can be reproduced by a reciprocal lattice volume element interacting with the remaining effects. For the SR case with a horizontal rotation axis, Greenough and Helliwell (1982b) first found the reflecting range due to the beam cross fire and the correlated spectral dispersion, modelled this with a suitable reciprocal lattice volume element, and then considered the interaction with the conventional component of spectral dispersion.

The surface containing the centres of the Ewald spheres which are present (before the conventional term is considered) can be approximated by an ellipse with major and minor axes $\gamma_{HE}/2$ and $\gamma_V/2$ respectively, where $\gamma_{HE} = (4\delta_{corr}^2 + \gamma_H^2)^{1/2}$ and $\delta_{corr} = \frac{1}{2}(\delta/E)_{corr}$ (figure 2.7). Figure 2.7 clearly shows how the left/right asymmetry of reflecting range, partiality, spot size, and blind region, arises. Once sample rocking width and conventional-source-type terms are accounted for, an expression for the angular reflecting range is obtained:

$$|\theta_R| = (L^2(\delta_{corr}^2 + \zeta \gamma_H)^2 + \gamma_V^2)^{1/2} + 2L\epsilon_n$$

where L is now the Lorentz factor, ζ the reciprocal space

horizontal coordinate and ϵ_h is given by

$$\epsilon_h = \frac{d^* \cos \theta}{\lambda} \left(\eta + \left(\frac{\delta \lambda}{\lambda} \right)_{\text{conv}} \tan \theta \right)$$

To decide whether a reflection is partial or not the sample rocking width Δ , or a spherical reciprocal lattice volume of radius E , can be compared with a unit Ewald sphere, and

$$\Delta = 2E/(d^* \cos \theta)$$

$$E = \frac{1}{L} \left[(\delta_{\text{corr}} d^* + \zeta \gamma_H)^2 + \gamma_V^2 \right]^{1/2} + \epsilon_h$$

The expression for θ_R clearly predicts a horizontal asymmetry due to the presence of the term in ζ . Reflections placed identically, except for the sense of their vertical displacement, will have the same reflecting range, while a change of sign in the horizontal displacement will give different results. For the source-monochromator-sample configuration at the SRS, the left-of-film reflections have increased reflecting range. At the Guinier position $\delta_{\text{corr}} = 0$, the term in ζ is squared, and the asymmetry is removed. The equations then reduce to those for a conventional source with an asymmetric cross-fire (Greenhough and Helliwell (1982a)). Greenhough and Helliwell (1982b) present two oscillation photographs of native glyceraldehyde 3-phosphate dehydrogenase, one taken at the Guinier position, the other with an overbent monochromator. In the latter, the spot extension caused by the increased reflecting range can clearly be seen for left-of-film reflections.

The asymmetric reflecting range also leads to an

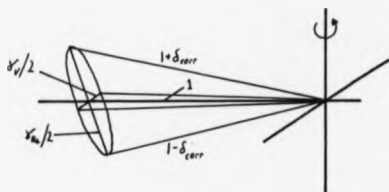


Figure 2.7 The correlation ellipse for a horizontal rotation axis, showing the Ewald sphere centres due to the beam cross-fire and direction-correlated wavelength.

asymmetric blind region and this is of great importance when processing oscillation data. Many reflections will become partial and many partials will be recorded on more than two films, such is the increase in size of the left-of-film blind region.

2.1.4 Diffraction spot size, shape, and energy profile

The range of wavelengths and ray directions which are diffracted by the sample to a single point on the film can be found, and so the energy spread, mean energy, and hence the resolution can also be found. By finding the film coordinates (H, V) of all the diffracted rays, the diffraction spots can be mapped in terms of position and energy. In this process the effects of crystal size, shape, and mosaicity must be taken into consideration. The first step is to consider those rays at the sample which pass through an infinitely small pre-crystal pinhole at the focus, with zero vertical cross fire. By temporarily neglecting the effect of the monochromator rocking width the diffraction spot produced by the direction correlated wavelength component alone can be mapped out. This turns out to be an inclined line on the film (since $\partial V / \partial H$ is constant for a particular point in reciprocal space), varying in length and inclination over the film, and which contains perfectly resolved energies with a linear relationship between position and energy (since $\partial H = \partial \delta_{\text{out}}$). Next, expressions are obtained for the spread in film coordinates for a particular energy due to the sample mosaicity (η), sample thickness (t), and

monochromator rocking width. Then the horizontal slit width (L_H) (assuming the sample can accept at least the whole of the central correlated fan of rays) is considered along with the associated focus energy gradient. If the energy gradient were zero, the slit would cause each energy to be smeared out over a horizontal distance equal to the slit width. Finally, the vertical beam cross-fire and vertical slit or sample size (L_V) (whichever is the smaller) are considered. The effect of the vertical cross-fire is, in most cases, very small compared to the simple vertical convolution due to the slit or sample height. This leads to the important result that, in finding the energy resolution at any point within a spot, only the horizontal film coordinate need be considered.

The total spread in the horizontal film coordinate due to the correlated component of energy alone is

$$(\Delta H)_{\text{corr}} = Y_H \left[\delta + \frac{C(\cos 2\theta + \epsilon')}{\cos^2 2\theta} \right] + \delta_{\text{corr}} \frac{Cc(2+d\epsilon')}{\cos^2 2\theta}$$

where S is the crystal to slit distance and C is the crystal to film distance. A given energy in the correlated component (whose general relative value is $\delta = (E - E_{\text{mean}})/E_{\text{mean}}$) is spread out over a horizontal distance ΔH^E on the film, given by the sum of the energy smearing factors, and the resolution of the correlated components is thus degraded to give the energy resolution at each point within the diffraction spot as

$$\Delta \delta = (\Delta H^E / (\Delta H)_{\text{corr}}) \Delta \delta_{\text{corr}}$$

Separating the energy smearing factors gives

$$\Delta \delta = \frac{\Delta \delta_{\text{corr}}}{|(\Delta H)_{\text{corr}}|} [|(\Delta H)_\eta| + |(\Delta H)_\zeta| + |(\Delta H)_\omega| + |(\Delta H)_{\text{foc}}| + |(\Delta H)_{Y_V}|],$$

and after substituting and rearranging

$$\Delta d = \frac{2\delta_{\text{corr}}}{(2H)_{\text{corr}}} \left[\frac{n\epsilon C + t\epsilon}{\cos 2\theta} + l_H + \frac{Y_V C\epsilon}{L \cos^2 2\theta} + Y_H \left(S + \frac{C(\cos 2\theta + \epsilon')}{\cos^2 2\theta} \right) \right] \left(\frac{\delta_\omega}{\delta_{\text{corr}}} - \frac{\delta_{\text{foc}}}{\delta_{\text{corr}}} \right)$$

where $\delta_\omega = \frac{1}{2} \left(\frac{\delta \lambda}{\lambda} \right) \omega = \frac{1}{2} \omega_{\text{diff}} \cot \theta$ and $\delta_{\text{foc}} = \frac{1}{2} \left(\frac{\delta \lambda}{\lambda} \right) l_{\text{oc}}$. The dimensions of the diffraction spot are given by

$$H_{\text{tot}} = Y_H \left[S + \frac{C(\cos 2\theta + \epsilon')}{\cos^2 2\theta} \right] + \left[\frac{n\epsilon C + t\epsilon}{\cos 2\theta} \right] + l_H \\ + \left[\frac{Y_H C\epsilon}{L \cos^2 2\theta} + \frac{(2t\epsilon')}{C \cos^2 2\theta} \right] [\zeta \delta_{\text{corr}} + \zeta \delta_{\text{foc}} + | \zeta \delta_\omega |]$$

and, for a symmetric l_V ,

$$V_{\text{tot}} = Y_V \left[S + \frac{C(1 - \zeta')}{\cos^2 2\theta} \right] + \frac{|n\epsilon C L \epsilon' \zeta| + t}{L \cos 2\theta} + l_V \\ + \left[\frac{Y_H C L \zeta (d^{*2} - 2\zeta')}{2 \cos^2 2\theta} + \frac{C L (2\zeta' - d^{*2} \zeta')}{\cos^2 2\theta} \right] (\delta_{\text{corr}} + \delta_\omega + \delta_{\text{foc}})$$

The diffraction spot is an inclined strip of vertical

width $V_{\text{uncorr}} = V_{\text{tot}} - 2V_{\text{corr}}$ and is symmetric about $V_c = C/L \cos 2\theta$, $H_c = C\zeta/\cos 2\theta$, (Wonacott (1977)). As the terms in $2H_{\text{corr}}$ and $2V_{\text{corr}}$ can be of the same or opposite sign, spot size and inclination vary over the film with large left/right differences.

The expressions above are of general use, even at the Guinier position, allowing the calculation of individual integration box shapes and sizes based on experimental conditions. Conversely, refinement of spot sizes will give a good estimate of sample mosaic spread (and crystal thickness t , if not carefully measured) assuming that machine parameters are accurately known. The effectiveness of these expressions is demonstrated

by the results described in §2.2.2.

That the energy-direction correlation present at the focus is preserved in each and every diffraction spot is potentially of great importance to the methods of optimized anomalous scattering. By centring the correlated energy spread on the absorption edge of an anomalous scatterer with, say, a metalloprotein, thus optimizing f' and f'' , the real and imaginary components of anomalous dispersion, sufficient information can be recorded that, in principle, it should be possible to phase the reflections without the use of multiple isomorphous replacement or multiple data sets individually collected at different wavelengths.

2.2 Experimental

2.2.1 Step Scan of Dy_2O_3 L_{III} Edge

In §2.1.1 and §2.1.2 it was noted that, at the Guinier position where there is no correlated contribution, the energy spread at the focus is dominated by the energy gradient term $(\lambda/p)\cot\theta$ due to the finite source size. By placing a small horizontal slit (width x_H) at the focus this term becomes $(x_H/p)\cot\theta$ and so can be reduced indefinitely at the expense of photon flux.

To demonstrate the resolution attainable the transmission (photons. sec^{-1}) through a thin Dy_2O_3 powder sample was measured on an arbitrary scale at the L_{III} absorption edge (1.5903 Å) by a simple ion chamber placed behind the sample. For this wavelength and a Ge (III) monochromator with an

oblique cut of 10.44° , the Bragg angle θ is 14.086° , the compression ratio A^2 is 0.1532 and the Guinier position is at $p_g = 3.207$ m. A $100 \mu\text{m}$ wide lead-jawed slit was placed in the focal plane and stepped horizontally across the focus in $50 \mu\text{m}$ intervals, thereby sampling a slightly different mean wavelength at each step. The measurements were made firstly without the sample and then with. The SRS was operating at 1.9 GeV with currents of 100 mA at injection and the data were corrected for the beam decay.

The scan without sample (I_0) (fig. 2.8a) gives a distance of approximately 3mm across the full focus and an FWHM value of 1.65 ± 0.05 mm. The corresponding $(\delta\lambda)_{\text{foc}}$ values are then 29eV and 16eV respectively. The energy resolution is given by the sum of the energy spreads due to the slit width ($\sim 1/1.65 \times 16 \text{ eV} \approx 1 \text{ eV}$), δa (1.2 eV), depth of focus (3 eV), and the monochromator rocking width ($\omega_{\text{diff}} \cot \theta = 2.4 \text{ eV}$) and therefore totals 7.6 eV, corresponding to a $\delta\lambda/\lambda$ of 9×10^{-4} . The scan with sample is shown in fig. 2.8b and the absorption ($= \ln(I/I_0)$) in fig. 2.8c. The Dy 'white line' feature evident in the absorption curve has a width of approximately 13 eV.

The experiment shows the resolution achievable at the Guinier position to be perfectly adequate for the sampling of absorption edge fine structure so that the wavelength may be accurately set for single wavelength optimised anomalous dispersion experiments.

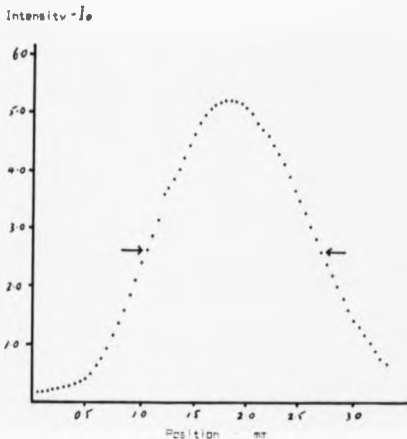


Figure 2.8a Step scan across the focus with no sample. The horizontal width of the beam at this wavelength (1.59 \AA) is $1.65 (0.05) \text{ mm FWHM}$. This corresponds to a source size of approximately 10.8 mm .

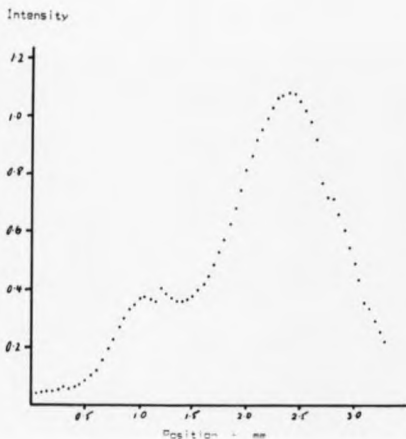


Figure 2.8b Scan across the Dy absorption edge with the powder sample in position at the focus.

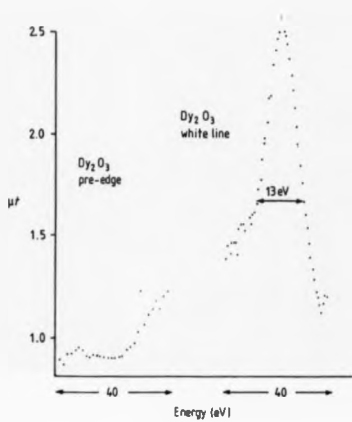


Figure 2.8c Absorption spectrum of Dy_2O_3 slightly below (left), and at (right), the $\text{Dy } L_{III}$ edge.

2.2.2 Polychromatic diffraction by a Re complex

Having established the possibility for a diffraction pattern in which each spot is drawn out into a streak along which the wavelength varies, a suitable sample was required in order to evaluate the method and the theory described above (§2.1). Crystals of heptahydrido bis (diisopropylphenol) phosphine rhenium ($\text{ReH}_7(\text{P}[\text{C}_3\text{H}_7]_2[\text{C}_6\text{H}_5])_2$), supplied by J.A.K. Howard, were chosen because of a) the high proportion of anomalous scatterer - one Re atom in 580 daltons, b) their uniform thickness (0.1 mm), and because c) the space group ($\text{P}2_1/\text{n}$) is centrosymmetric, all phases are 0 or π and the variation in f'' has a negligible effect on the reflected intensity. The crystal structure has been solved previously (Howard et.al. (1983)) from low temperature X-ray data collected on a Nicolet P3/m 4 circle diffractometer, and the refined cell parameters were found to be $a = 11.271 \text{ \AA}$, $b = 13.405 \text{ \AA}$, $c = 17.017 \text{ \AA}$ and $\beta = 95.53^\circ$.

The aim of the experiment, therefore, was to observe the effects of f' on the reflection profile at the rhenium L_{III} absorption edge (1.1772 \AA). No modification of the standard SRS protein crystallography station 7.2 was required apart from the introduction of an adjustable horizontal slit to constrain the energy resolution within the reflection profile. The slit was set at $70 \mu\text{m}$ and positioned at the focus with the sample beyond so that the non-diverging beam flooded most of the crystal, which therefore accepted all the correlated energy bandwidth available (figure 2.9).

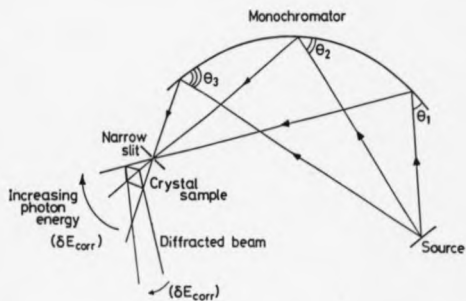


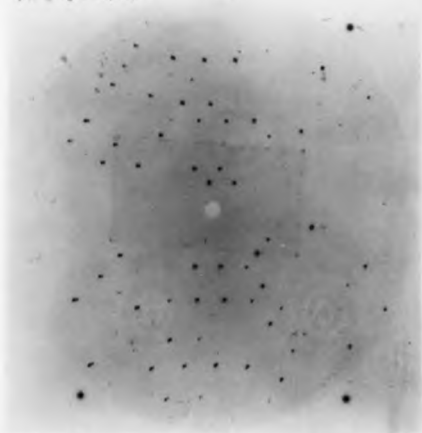
Figure 2.9 Schematic of the polychromatic profile experiment in which the monochromator is 'overbent' with respect to the Guinier position.

A Si(111) monochromator with an asymmetric cut of $6.73 \pm 0.05^\circ$ (Helliwell, unpublished data) was used; the slit to crystal distance was 100 mm and the crystal to film distance was 54.5 mm, corresponding to a resolution limit on a flat plate cassette of 1.5 Å. The monochromator was overbent so that the focal plane was at $p' = 2.4$ m, which leads to calculated value for δE_{corr} of 85 eV. However, later analysis of direct beam exposures indicated that only 156 mm of the available 200 mm was illuminated, consequently $\delta E_{\text{corr}} = 67$ eV (Greenhough, pers. comm.).

The wavelength was first calibrated by measuring the intensity transmitted by a crystal, similar to that used in the diffraction experiments, as the monochromator was stepped in intervals of 0.01° . Two sequences of photographs were then recorded, on CEA Reflex 25 film, with the wavelength spread centred in the first sequence, on the point of inflection of the absorption curve, and in the second, 200 eV to the low energy side. Each sequence of exposures consisted of: a still at 0° spindle angle; a still at 85° ; an oscillation photograph from 85° to 89.95° ; and finally a still at 89.95° . The sample was mounted such that the x-ray beam was perpendicular to the large face of the thin crystal at a spindle angle of 90° .

The two oscillation photographs (referred to as 'on-edge' and 'off-edge') are shown in figure 2.10. The 'on-edge' photograph shows apparent uniform fine structure in each and every reflection, as would be expected for a small heavy atom complex where the absorption profile of

442 05-5495



15-4

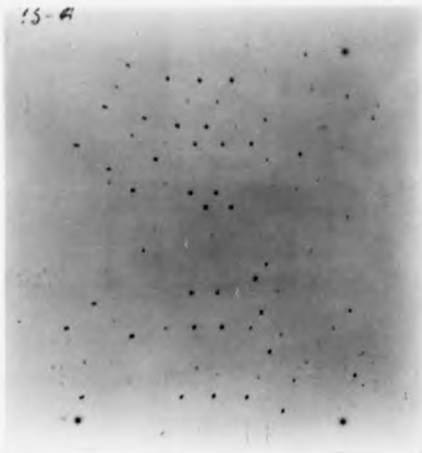


Figure 2.10 The 'on-edge' (top) and 'off-edge' photographs.

the heavy atom dominates the diffracted intensities. Such fine structure is not observed in the 'off-edge' photograph and cannot be attributed to a split sample since the 'on-edge' sequence was recorded first.

The spots recorded on the still photographs were used with the program OSCAR (see §4.4) to refine the sample orientation and index the reflections on the oscillation photographs. With only 30 partial reflections available the refined orientation is relatively imprecise; however, with a small unit cell the mis-indexing of reflections is highly unlikely. The large oscillation range leads to the same reflections being stimulated at the two (slightly) different wavelengths, although their position on each film is slightly altered. The still photographs were also used to check that any reflection chosen for analysis was indeed fully recorded on the oscillation photographs.

With the atomic coordinates of the crystal structure available it was possible to calculate the individual contributions to the structure factors of all the fully recorded low angle reflections. Two such reflections were selected, those with Miller indices 020 and $\bar{1}20$, and these are stimulated at very similar Bragg angles, so that their absorption profiles, as a function of wavelength, should also be very similar. The Re atom makes a small contribution to the $\bar{1}20$ structure factor ($F_{Re} = 36$, $F_{tot} = -110$), but it dominates that of the 020 ($F_{Re} = 259$, $F_{tot} = 134$). It was therefore expected that the variation in the Re f' across

the absorption edge would modify the intensity profile of only the latter reflection, and that division by the former reflection profile would eliminate the dominating absorption effect and reveal the variation due to f' .

The reflections of interest were digitised on a Scandig flat bed densitometer using an aperture of $10 \times 100 \mu\text{m}$ with a step size of $10 \mu\text{m}$. The scan direction was along the streak axis, which was very close to horizontal for these low angle reflections. The four central scans were averaged together, making the raster effectively $10 \times 400 \mu\text{m}$, and the scans above and below the spots were used to provide an average local background which was then subtracted from the reflection profile. The uncorrected profiles for the two reflections at the two different wavelengths are shown in figure 2.11. Since the two reflections on each film were not of exactly the same length (differing by up to $50 \mu\text{m}$ in 1.5 mm) along the streak axis, the longer was compressed using linear interpolation to match the shorter. To achieve this the start and end point of each profile was defined as being where the profile intensity reached twice the average background intensity.

Once these corrections had been applied the 020 profile was divided by the $\bar{1}20$ profile at each wavelength. The results of this division are plotted in figure 2.12, which clearly shows a large variation at the 'on-edge' wavelength, whilst 'off-edge' the effect is absent. To test the effect of misalignment of the reflections, one

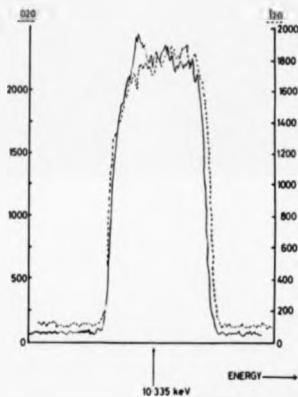
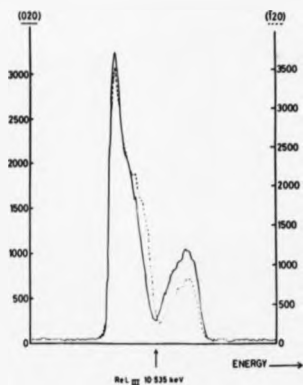


Figure 2.11 The uncorrected densitometer profiles of the two reflections in the 'on-edge' (top) and 'off-edge' cases.

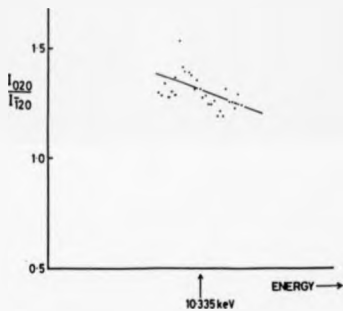
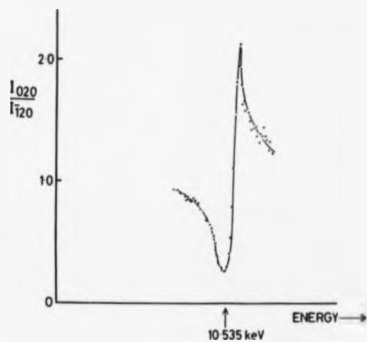


Figure 2.12 The divided intensity profiles (020/120) in the 'on-edge' (top) and 'off-edge' cases.

profile was slipped by up to 4 raster points (i.e. 40 μm) in either direction with respect to the other. Although this produced changes in the magnitude of the peaks in the 'on-edge' division, the shape of the plot remained substantially unaltered. Clearly, any estimate of f' should be based on a least squares analysis of many reflections. However, Helliwell (unpublished calculation) has estimated that the minimum in the 'on-edge' divided profile of figure 2.12a corresponds to a f' of -27 e^- which compares well with values obtained for Pr and Sm by Templeton et.al. (1980).

A detailed analysis of the energy resolution at any point in the reflection profile has been given by Greenhough, Helliwell and Rule (1983) (see appendix II). Taking into account sample mosaic spread, sample dimensions, size of slit and slit-sample-film distances, as well as beam geometry, a value for δE of approximately 11 eV (or $\delta E/E = 1 \times 10^{-3}$) was obtained for the low angle reflections considered here. These authors have also compared the spot shapes recorded in this experiment with those predicted by the theory outlined earlier in this chapter (2.1.4) and these show very good agreement.

A brief attempt to extend this experimental approach to a protein sample has been made by Helliwell and Greenhough (pers. comm.) using a crystal of Manganese/Calcium Pea Lectin. However, they concluded that the anomalous scattering signal of the Mn K edge for 2 Mn atoms in 50000 daltons was too weak to be observed in their experiment.

In general, the metal L series would be superior especially if absorption edge white line effects were accessible. Despite this, it is in the phase determination of reflections from protein crystals containing a single anomalous scatterer in the asymmetric unit, that the polychromatic method is perhaps best suited. Changes in sample absorption across the absorption edge would be very much smaller than in the present case, and, at high angle where the atomic scattering factor falls off rapidly, the anomalous signal would be comparatively large.

There is potentially an enormous amount of information generated by a polychromatic diffraction experiment with a protein sample, and, aside from the experimental considerations, much thought and effort will be required to establish the means by which this information can be extracted and interpreted. Alternatively the aim might be more modest; perhaps to accurately phase a subset of reflections, which could then be incorporated into direct methods phasing of protein data (e.g. Bricogne (1984)). One aspect of the experiment which is open to improvement is that of the detector. For example, the FAST TV diffractometer, with good sensitivity and spatial resolution, could be placed a comparatively long way from the sample so as to spread the diffraction streaks across many pixels. Alternatively, by using a small spindle rotation step angle, parts of a reflection could be spread over more than one image, thereby allowing three dimensional profiles to be recorded. The 'profile' would be the energy axis and the energy

resolution along the 'streak' would be set by the step angle. Use of on-line prediction software would greatly reduce the volume of data, and, of course, the micro-densitometer stage would be eliminated completely.

Despite the fact that there is still a long way to go before the polychromatic method could be routinely applied to suitable samples, the experiment described here does represent a successful first step towards that goal.

CHAPTER THREE

PROPERTIES AND SIGNIFICANCE OF PURINE NUCLEOSIDE PHOSPHORYLASE

Nucleoside phosphorylases were first recognised as a distinct group of enzymes through the studies of Kalcker (1945a, b). Previously, it had been believed that the enzymic cleavage of the glycosidic bond of nucleosides was the result of direct hydrolysis. Only after it had been shown that enzymic cleavage of nucleosides is stimulated by phosphate or arsenate (Klein (1935)) and that glycogen phosphorylase catalyzes the depolymerization of glycogen by the liberation, not of glucose, but of glucose-1-phosphate (Cori et.al. (1937)), was it suggested that phosphorolysis rather than hydrolysis might be the mechanism involved. The next step was the identification and isolation of the reaction product ribose-1-phosphate (R-1-P) (Kalcker (1947)), followed by the discoveries that phosphorolysis of purine deoxyribonucleosides yielded deoxy-R-1-P (dR-1-P), and that incubation of R-1-P or dR-1-P with an appropriate purine base resulted in the synthesis of a purine nucleoside (Friedkin and Kalcker (1950), Friedkin (1950)).

However, it is the relatively recent appreciation of the importance of purine nucleoside phosphorylase (PNP) in immunodevelopment and purine nucleoside analog metabolism that has stimulated detailed structural

and kinetic studies and has generated strong interest in developing PNP inhibitors and nucleoside analogs resistant to cleavage.

3.1 The function of PNP

Mammalian purine nucleoside phosphorylase (PNP) (nucleoside phosphorylase; purine nucleoside: orthophosphate ribosyltransferase; EC 2.4.2.1) catalyzes the reversible phosphorolysis of ribonucleoside and 2'- deoxyribonucleoside derivatives of hypoxanthine, xanthine, guanine and many of their analogs (figure 3.1):



Although in man and other species the synthesis of nucleosides is greatly favoured under equilibrium conditions, PNP acts primarily in the phosphorolytic direction in intact cells. Figure 3.2 shows how PNP functions as a catabolic enzyme when coupled with guanase and xanthine oxidase and as a salvage enzyme by providing substrates for hypoxanthine-guanine phosphoribosyltransferase (HGPRT). The liberated sugar phosphates may enter the pathway of carbohydrate metabolism or may participate in nucleoside exchange reactions where PNP couples with a pyrimidine nucleoside phosphorylase (Parks and Agarwal (1972)). R-1-P, via its conversion to R-5-P by phosphoribomutase, is also an important precursor of 5-phosphoribosyl-1-pyrophosphate (PRPP).

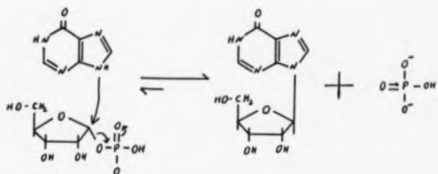


Figure 3.1 The PNP reaction. In this example PNP catalyses the synthesis of inosine from hypoxanthine and ribose-1-phosphate.

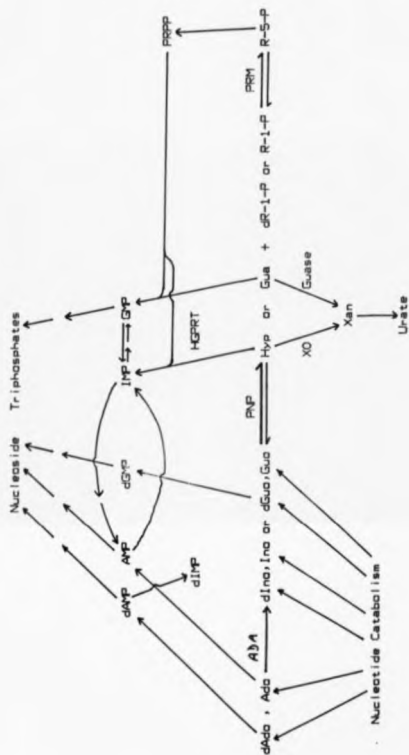


Figure 3.2 Purine salvage and catabolism in the pathways around PNP.

Parks et.al. (1975) have shown that guanosine, a R-I-P donor, is incorporated much more rapidly into the nucleotide pools of human erythrocytes than guanine. It is also important to note that both ribo- and deoxyribonucleosides are salvaged by the PNP pathway to form only ribonucleotides. Thus, coupled with adenosine deaminase (ADA) PNP may provide a mechanism for limiting the production of deoxyribonucleotides to their synthesis by ribonucleoside diphosphate reductase, a highly regulated enzyme (Moore and Hurlbert (1966)).

PNP shows little regulation but is present in very high activities per mg of protein in man, particularly in extracts of kidney, peripheral lymphocytes and granulocytes. Erythrocytes, which are deficient in purine biosynthesis de novo, have about the same activity per ml of packed cells as peripheral lymphocytes (Parks et.al (1975)) and are the richest source of human PNP. Turner et.al (1974) have noted that reticulocytes have almost 7 times as much activity as mature erythrocytes. In contrast to lymphocytes, erythrocytes transport nucleosides into the cell as rapidly as they can be split by PNP (Tax and Veerkamp (1978)). Parks et.al. (1975) have estimated a hypothetical rate in excess of 500g inosine cleaved per hour per 2.5 l of erythrocytes illustrating the formidable phosphorolytic capacity of human blood. It is also important when extrapolating from laboratory animals to man to note that PNP activity in erythrocytes is 5-fold lower in mice (Burgess et.al.,

unpublished results), 15-fold lower in rats (Agarwal et.al. (1975)), and extremely low or nil in dogs and cats (Parks and Agarwal (1972)).

3.2 Significance of PNP in chemotherapy

3.2.1 Relationship to immunodeficiency

In 1972 a deficiency of ADA was found in two unrelated patients with severe combined immunodeficiency disease (Giblett et.al. (1972)) and this was followed by the report of a child, exhibiting lymphopenia, severe T cell dysfunction, but normal B cell function, who was completely lacking in PNP activity (Giblett et.al. (1975)). Ammann (1978) has since detailed the clinical manifestations of the immunological aberrations in PNP deficient patients. Peptide mapping of the abnormal enzyme subunit from the families of deficient patients has confirmed that the mutation is in the structural gene for PNP (Gudas et.al. (1978b), McRoberts and Martin (1980)).

Studies of incompletely deficient patients suggest that proliferation-dependent T-suppressor cells are impaired while proliferation-independent T-helper cell functions are retained (Gelfand et.al. (1978)), which may account for the exaggerated humoral response often observed (Ammann (1978)). These findings suggest a means of selectively suppressing cellular immunity for therapeutic purposes. An effective PNP inhibitor might be useful in treating T-cell leukemias (Stoeckler et.al.

(1980)), suppressing the host vs. graft response (Parks et.al (1981), or in countering autoimmune disease (Kazmers et.al. (1981)) without destroying the patients humoral immunity. In addition xanthinuria and xanthine gout might be treated by a PNP inhibitor which would shift the purine excretion pattern toward the more soluble nucleosides. An inhibitor might also be useful in the treatment of secondary gout due to rapid cytolysis after irradiation or chemotherapy (Parks et.al (1981)).

Since there is no evidence that missing products of the PNP reaction are detrimental to the immune system, attention has been directed toward the potential toxicity of accumulated substrates. Of the substrates only deoxyguanosine (dGuo) can be phosphorylated and in the absence of PNP the other substrates guanosine, inosine and deoxyinosine are excreted in elevated quantities in urine (Cohen et.al (1976), Siegenbeek van Heukelom et.al. (1977)). Deoxyguanosine alone of the substrates inhibits growth of the S49 mouse T-cell lymphoma line at 10^{-5} M concentrations (Chan (1978), Gudas et.al. (1978a)). Cells with a mutant ribonucleoside diphosphate reductase that is not subject to allosteric control by deoxyguanosine triphosphate (dGTP), show resistance to dGuo toxicity (Ullman et.al. (1979)). The accepted model for explaining the effects of PNP deficiency is that shown in figure 3.3. In the absence of PNP (or in the presence of an effective inhibitor of PNP) dGuo accumulates and is phosphorylated to form dGTP thereby inhibiting the

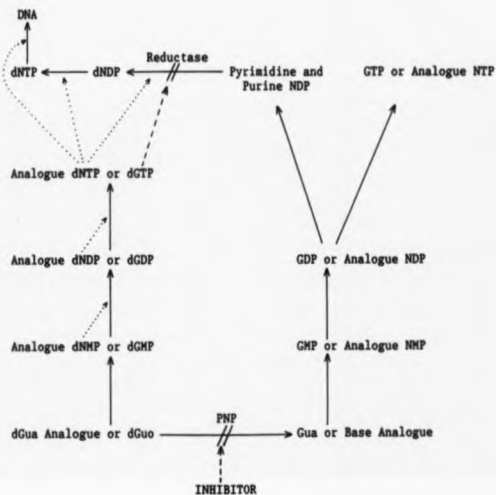


Figure 3.3 Proposed mechanism for the toxicity of PNP inhibitors and deoxyguanosine analogues resistant to phosphorolysis. NMP, NDP and NTP are nucleoside mono-, di-, and tri-phosphates. (---> //) are known sites of inhibition, and (----->) are other potential sites.

reductase causing a depletion of deoxycytidine triphosphate (dCTP) and so blocking the synthesis of DNA. Carson et.al. (1979) suggest that the T-cell specificity is due to the selective trapping of dGuo nucleotides by the relative activities of phosphorylating and dephosphorylating enzymes in these cells.

While the understanding of inborn immunodeficiency diseases has progressed enormously in recent years, there are still many unexplained clinical and experimental observations. PNP deficient patients often display neurological abnormalities and anaemia which are not attributable to autoimmune disease (Ullman (1978)). There is evidence of an association between haemolysis and perturbations of erythrocyte nucleotide levels (Valentine et.al. (1974)). Inhibition of S-adenosylhomocysteine hydrolase is known to occur secondarily in ADA (Hershfield et.al. (1979), PNP and HGPRT deficiency (Hershfield (1981)). The clinical usefulness of PNP inhibitors may ultimately be determined by the causes and importance of these and other findings.

3.2.2 Purine Analog Metabolism

There has also been much interest in the role of PNP in the metabolism of purine analogs used in chemotherapy. There are three main areas of study. Firstly, PNP cleaves the nucleosides of many guanine and hypoxanthine analogs to release the bases. It is well known that the ribonucleosides of some purine analogs used in tests of anti-tumour activity are as effective as their free bases

(Montgomery et.al. (1962)). 6-thioguanosine and 6-selenoguanosine are incorporated into the nucleotide pools of erythrocytes to a much greater extent than their free bases, probably because upon cleavage by PNP the PRPP precursor R-1-P is made available. (Parks et.al. (1975)) The rapid cleavage of the deoxynucleosides of guanine analogs by PNP raises the question of whether they would be directly phosphorylated (as with dGuo) if their phosphorolysis were inhibited. Investigations with mice suggest that 8-2'-deoxy-6-thioguanosine might be a more effective anticancer agent than 6-thioguanine (Lepage et.al.(1964)). Clinical tests proved negative although the higher activity of PNP in human blood may have prevented large enough doses of the nucleoside from reaching the target cells.

Secondly, although PNP functions predominantly as a phosphorolytic enzyme, the intracellular synthesis of analog nucleosides has been demonstrated. 1-ribosylallopurinol has been identified in the urine of patients receiving the drug allopurinol (Krenitsky et.al. (1967)). Studies with mice show that although the nucleoside is a product of the PNP reaction it is a poor substrate for PNP, resulting in the analog being trapped in its nucleoside form (Nishada et.al. (1979)). 6-thioguanosine is synthesized from 6-thioguanine in mouse Sarcoma 180 cells and it has been suggested that the PNP reaction lowers the potency of 6-thioguanine by making it unavailable for reaction with its target enzyme HGPRT.

Finally, the PNP reaction can synthesize analog sugar phosphates from sugar-modified nucleosides. Recent studies have shown that C(5') modified analogs of methylthioadenosine and C(5') substituted inosines are equally toxic in cells with high PNP activity (Parks et.al. (1981)). Tumor cell lines deficient in methylthioadenosine phosphorylase may be treated by toxic sugar phosphates generated by the PNP reaction (Kamatani and Carson (1980)).

Recent interest has centred on analogs of 2'-deoxyguanosine and 5' modified nucleosides in an effort to identify PNP resistant analogs that might be phosphorylated directly and form analogs of dGMP, dGDP and dGTP (Stoeckler et.al (1982a,b)). Ultimately the chemotherapeutic value of such analogs also depends on their capacity to be activated by nucleoside and nucleotide kinases, their susceptibility to inactivation by nucleotidases, and the biological activities of the analog nucleotides. If dGuo analogs which have good substrate activity with PNP are to be useful, a potent PNP inhibitor may also be required to prevent cleavage in the bloodstream since rapid phosphorolysis of dGuo is observed in erythrocytes even in the presence of 8-aminoguanine, one of the better PNP inhibitors identified to date (Stoeckler et.al. (1982a)).

3.3 Properties of PNP

3.3.1 Purification

PNP is widely distributed in nature and has been identified and studied in a number of tissues in mammals,

chickens, fish, yeast, and in several species of bacteria. These have been tabulated by Parks and Agarwal (1972). Crystalline human erythrocytic PNP was first isolated (Agarwal and Parks (1969)) by modification of procedures using only the gentle techniques of ion-exchange chromatography, ammonium sulphate fractionation and gel filtration, (Tsudoi and Hudson (1957), Abrams et.al. (1965)). Similar methods have since been used to purify the enzyme from rabbit liver (Lewis and Glantz (1976a)), brain (Lewis (1978)) and erythrocytes (Savage and Spencer (1977)), from bovine liver (Ikazawa et.al. (1978)) and brain (Lewis and Glantz (1976b)), from chicken liver (Murakami and Taushima (1975)), and from Chinese hamster liver, kidney and V79 tissue culture cells (Milman et.al. (1976)). Affinity chromatography, using inosine for gentle ligand displacement, has also been employed successfully in purifying PNP from human erythrocytes (Zannis et.al. (1978), Osbourne (1980)), as well as from cultured human fibroblasts (Zannis et.al. (1979)) and granulocytes (Wiginton et.al. (1980)).

The human erythrocytic PNP used in all the crystallographic studies described in the following chapters, has been kindly supplied by Drs. R.E. Parks Jr. and J. D. Stoeckler of Brown University, Providence, Rhode Island, and they have developed the following purification procedure (Stoeckler et.al. (1978b)).

1. PNP from pooled, outdated (21 days) human CPD blood (provided by the Rhode Island Blood Centre) undergoes large scale partial purification by the New England Enzyme Centre

and chromatography on a 10 x 60 cm calcium phosphate gel-cellulose column (Agarwal et.al. (1978)).

2. The enzyme is concentrated by precipitation with 65% ammonium sulphate and dialyzed against 0.025 M imidazole-HCl containing 1 mM dithiothreitol (DTT).

3. A 2.5 x 25 cm column of DEAE cellulose (acetate form) is equilibrated with 0.03 M Tris-acetate, pH 7.5, containing 1mM DTT. The dialyzed enzyme is absorbed on the column, washed with about 50 ml of the same buffer and eluted with a linear gradient (0.03 - 0.35 M : total volume 600 ml) of Tris-acetate, pH 7.5. About 95% of the enzyme emerges in the range of 0.06 - 0.20 M Tris-acetate, and the fractions containing enzyme activity are pooled.

4. Residual coloured proteins and ampholytes are removed by chromatography on a small (1.5 x 10 cm) calcium phosphate gel-cellulose column and then by gel filtration on Sephadex G-100.

5. Eluates from the preceding columns are concentrated in a Collodion Bag Apparatus and the final preparation, containing the multiple isoelectric variants, is recrystallized twice from ammonium sulphate.

Large scale purification of the enzyme has been facilitated by the advent of chromatofocusing which involves absorption of proteins to an ion exchange column at a pH above their isoelectric points and the step-wise elution at pH values slightly below their isoelectric points. The technique should make it possible to isolate quantitatively the different electrophoretic variants of PNP, and it has now

replaced the DEAE-cellulose chromatographic stage in the above procedure.

3.3.2 Subunit Organisation

Although trimeric enzyme structures are not common (Klotz et.al. (1975)), evidence from many sources has suggested that human PNP has three subunits of equal size: a fact now confirmed by the X-ray analysis at 6Å resolution (55·2). Subunit molecular weights of 30,000 to 33,000 daltons have been determined by SDS gel electrophoresis for the enzyme from red cells (Zannis et.al (1978), Osbourne (1980), Agarwal et.al. (1973) and Stoeckler et.al. (1978a)), leukemic granulocytes (Wiginton et.al. (1980)), and placenta (Ghangas and Rheem (1979)). The original crystalline human enzyme was shown to bind approximately three molecules of substrate per enzyme molecule (Agarwal and Parks (1969)).

In addition, the electrophoretic patterns of genetic variants and of hybrids formed between the enzymes from human foetal liver and mouse liver, suggest a trimeric protein (Edwards et.al. (1971)). Edwards et.al. (1971) also reported that PNP is the product of a single autosomal gene locus which was later assigned to chromosome 14 (Ricciuti and Ruddle (1973)). The PNP from Chinese hamster tissues (Milman et.al. (1976)), bovine spleen (Edwards et.al. (1973), liver (Ikazawa et.al. (1978)) and thyroid (Carlson and Fischer (1979)), and from rabbit erythrocytes and liver (Savage and Spencer (1977)) are all believed to be trimeric. Suggestions of a monomeric rabbit liver PNP (Lewis and Glantz (1976a)) and dimeric protein from human erythrocytes

(Lewis and Lowry (1979)), rabbit brain (Lewis (1978)) and bovine brain (Lewis and Glantz (1976b)) all seem to be at variance with the findings of other workers.

3.3.3 Electrophoretic Heterogeneity

Native human erythrocytic PNP shows considerable electrophoretic heterogeneity resulting from post translational modifications. Edwards et.al. (1971) have shown by starch gel electrophoresis that the enzyme from normal individuals migrates into at least seven bands with the greater activity in the anodal forms. Ten to fifteen variants were distinguishable in the patterns from individuals with rare mutant alleles. Furthermore, a correlation was found between the in vivo aging of erythrocytes and the progressive loss of alkaline variants, and appearance of acidic variants (Turner et.al. (1971)). Electrophoretic heterogeneity has also been detected by polyacrylamide gel isoelectric focusing of human PNP from a variety of tissues. For the erythrocytic enzyme isoelectric points ranging from 5.0 - 6.1 (Zannis et.al. (1978)), 5.0 - 6.4 (Zannis et.al. (1979)), 5.24 - 5.86 (Changas and Rheum (1979)) and 5.25 - 5.85 (Agarwal et.al. (1975)) have been reported.

Electrophoresis of the subunits of human erythrocytic PNP in 8.5 M urea has demonstrated the existence of four major components with isoelectric points between 6.20 and 6.63, and two minor ones (Zannis et.al. (1978)). Two additional alkaline subunits have been detected in the erythrocytic enzyme from the heterozygous parents of a

PNP-deficient patient (McRoberts and Martin (1980)). Zannis et.al. (1978) have proposed that the acidic subunits differ from the most alkaline one by 2, 3 and 4 negative charges, and so 12 differently charged trimers can be assembled. At least 10 variants have recently been observed using chromatofocusing techniques (Dowd et.al. (1983)).

3.3.4 Kinetics

An interesting aspect of the PNP reaction is activation by high substrate concentrations as indicated by downward deflections away from linearity of Lineweaver-Burk plots (Lineweaver and Burk (1934)). This effect has been noted for purine bases (Krenitsky (1967), Stoeckler et.al. (1982a) and their nucleosides (Edwards et.al. (1971), Ikazawa et.al. (1978)), as well as for phosphate (Agarwal and Parks (1969), Moyer and Fischer (1976)). For the human erythrocytic enzyme the deviation from linearity occurs at inosine concentrations of greater than 200 mM (Kim et.al. (1968)).

Turner et.al. (1971) have shown that substrate activation by inosine is a characteristic of the more acidic variants of human erythrocytic PNP and is not observed with the more alkaline isozymes from young erythrocytes. It therefore appears that the substrate activation is closely related to the progressive post translational modification of the enzyme. Loss of substrate activation can be effected by treatment with the sulfhydryl reagent 5,5'-dithiobis (2-nitrobenzoic acid) (DTNB) (Agarwal and Parks (1971)). Three moles of DTNB reacting with one mole of enzyme causes a 60% loss of enzyme activity and a 4- to 5-fold increase in

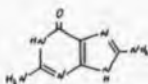
K_m value for inosine. Dithiothreitol reverses these kinetic changes which are not protected against by guanine.

The pH dependence of kinetic parameters suggest the participation of both cysteine (pK_a 8.2-8.5) and histidine (pK_a 5.5-6.4) in the catalytic mechanism of PNP (Agarwal and Parks (1964), Carlson and Fischer (1979)). Cysteine is further implicated by the complete inactivation of the enzyme by 4 moles of p-chloromercuribenzoate (PCMB) per mol of PNP (Agarwal and Parks (1971)), against which formycin B offers protection (Carlson and Fischer (1979)). An essential arginine residue has also been identified and this is partially protected by R-1-P, phosphate, arsenate or inosine (Jordan and Wu (1978a)).

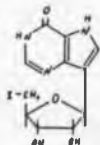
Kinetic studies suggest an ordered reaction mechanism for PNP. Carlson and Fischer (1979) and Moyer and Fischer (1976) report that the phosphate is the first substrate to bind to bovine thyroid PNP, and R-1-P is the last product released. Conversely the calf spleen (Krenitsky (1967)), and human erythrocytic enzymes (Kim et.al. (1968)) are thought to bind the nucleoside or purine base first and release it last.

3.4 Inhibitors of PNP

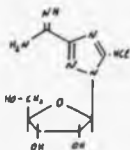
In recent years studies of PNP have concentrated on the identification of useful inhibitors. Figure 3.4 shows the best inhibitors characterized to date. 8-aminoguanine (8-AG) has an apparent K_i value of 2×10^{-7} M for the human erythrocytic enzyme, compared to the value for the natural substrate, guanine, of 5mM (Stoeckler et.al. (1982a)).



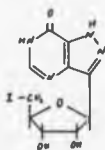
a.



b.



c.



d.

Figure 3.4

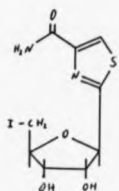
PNP inhibitors.

- a) 6-aminoguanine $K_i < 2 \times 10^{-7}$ M
- b) 9-deaza-5'-iodo-inosine $K_i < 10^{-6}$ M
- c) TCM $K_i = 5 \times 10^{-6}$ M
- d) 5'-iodoformycin B $K_i = 7 \times 10^{-6}$ M

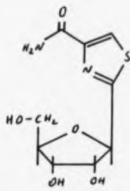
8-AG is an alternative substrate of PNP and 8-aminoguanosine (8-AGuo) can be synthesized enzymatically with R-1-P. 8-AGuo has a K_i value of 17mM and parallels the observations with the natural substrates that the nucleosides have a lower affinity for the erythrocytic enzyme than the purine bases (Krenitsky et.al. (1968), Zimmerman et.al (1971), Jordan and Wu (1978b)). 8-AGuo can serve as a pro-drug from which the less soluble base can be generated intracellularly. The lack of substrate or inhibiting activity prohibits metabolic depletion of the inhibitor and may prevent secondary effects on purine metabolism. Various other C(8) substituents have been tested but all have a lower affinity for PNP than 8-AG.

The tightest binding nucleoside tested to date is 9 desza-5'-deoxy-5'-iodoinosine (I9DI). Initial results indicate that it is a competitive inhibitor with a K_i value in the 10^{-7} M range (Stoeckler et.al. (1983)). The iodo substituent at C(5') appears to be partly responsible for the tight binding as is also found with other inhibitors.

TCNR is an unlikely looking PNP inhibitor as it only has one heterocyclic ring. It was synthesized and tested as an antiviral agent by Witkowski et.al. (1973). TCNR is resistant to phosphorolysis but it undergoes adenosine kinase dependant phosphorylation and so is not a 'pure' PNP inhibitor (Willis et.al. (1980)). Two similar structures (figure 3.5), 2-B-D-ribofuranosylthiazole-4-carboxamide and its 5'-deoxy-5'-iodo derivative, whilst not being good inhibitors themselves, do illustrate the enhancement of



a.



b.

Figure 3.5 These thiazole nucleoside inhibitors of PNP demonstrate the enhanced affinity afforded by the c(5') iodine substituent.

a) 2-(5'-deoxy-5'-iodo- β -D-ribofuranosyl)thiazole-4-carboxamide,

$$K_1 = 5 \times 10^{-5} \text{ M}$$

b) 2- β -D-ribofuranosylthiazole-4-carboxamide, $K_1 = 5 \times 10^{-4}$ M

2- β -D-ribofuranosylthiazole-4-carboxamide, $K_i = 5 \times 10^{-4}$ M

affinity afforded by the halogen at the C(5') position. This suggests that a similar iodine substitution might improve the affinity of TCNR for PNP as well as eliminate the possibility of phosphorylation and secondary inhibitory effects.

5'-deoxy-5'-iodoformycin also demonstrates the improvement in binding due to iodination at C(5'). Formycin B, an antibiotic (Koyama and Umezawa (1965)), inhibits erythrocytic PNP with a K_i value of 10^{-6} M (Townsend et.al. (1978)), whereas the iodinated derivative shows a greater than 10-fold improvement (Stoeckler et.al. (1982a)).

Recently, a new, relatively potent inhibitor, acyclovir (figure 3.6), has been reported (Tuttle and Krenitsky (1984)). This is a clinically useful antiherpetic agent (Elion et.al. (1977), Schaeffer et.al (1978)). In herpes infected cells and, to a lesser extent, in uninfected cells acyclovir is anabolised to its mono, di-, tri-phosphate metabolites. The major factor in the antiherpetic activity appears to be the inhibitory effects of the triphosphate on viral DNA replication. The 8-hydroxy metabolite has a K_i value of 4-7 mM, but it is the phosphorylated metabolites that show a significant improvement over their guanine nucleotide counterparts. At high phosphate concentrations (50 mM) the K_i values for the mono (120 μ M), di- (0.51 μ M) and tri (43 μ M)-phosphate esters of acyclovir are 1/2, 1/1200 and 1/26 of those for dCMP, dGDP and dGTP respectively. At physiological phosphate concentrations (\sim 1 mM) the K_i values

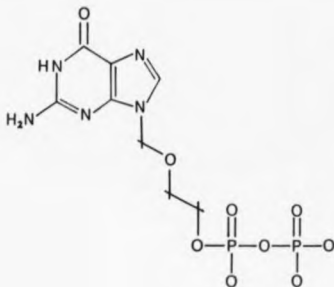


Figure 3.6 Acyclovir diphosphate is the most potent inhibitor of PNP yet reported, having a K_i value of $0.0087 \mu\text{M}$ at low phosphate concentrations (Tuttle and Krenitsky (1984)). That the mono- and triphosphates are inferior inhibitors suggests the theory that the second phosphate group binds in the phosphate binding site of PNP.

for these phosphorylated metabolites are reduced to 6.6, 0.0087 and 0.31 μ M respectively. Hence, acyclovir is the most potent PNP inhibitor yet reported.

Many other analogs have been tested for inhibition of PNP but none show an improvement over those described above, although some general features are apparent. Modifications at C(2') and C(3') drastically reduce PNP inhibition yet there is surprising flexibility at the C(5') position. With the exception of 5' modified nucleosides and some 8-azapurines, ring modified bases and nucleosides bind poorly or are unreactive with erythrocytic PNP. In addition, all methyl or thio substitutions on the purine ring reduce the affinity for the enzyme (Krenitsky et.al (1968)).

Although selective immunosuppression has been induced in mice by administration of dGuo alone (Doach et.al. (1980)), in the absence of exogenous dGuo the 'pure' PNP inhibitors available to date are generally non toxic to various mouse and human cell lines. Growth inhibition is seen, however, with inhibitors that have additional sites of action. Formycin B inhibits growth of human B and T cell lines (Willis et.al. (1980), Willemot et.al. (1979)) and peripheral lymphocytes (Osbourne et.al. (1980)). The growth inhibition is associated with inhibition of DNA synthesis in peripheral lymphocytes, but no synergy with dGuo toxicity is observed (Cowan et.al. (1981)). Growth inhibition by TCNR is related to inhibition

of inosine monophosphate dehydrogenase in cultural lymphoblasts of B cell origin (Willis et.al. (1980)). PNP inhibition by 8-AGuo and 8-AG has been studied in a wide variety of cell types (Kazmers et.al. (1981), Stoeckler et.al. (1982a)), and it appears that cell types vary greatly in their response to PNP inhibitors as well as in their sensitivity to dGuo. In addition, little is known about the transport of PNP inhibitors and this may well differ among different cell lines, as may the transport of dGuo.

It would appear that more potent PNP inhibitors are needed since patients with less than 1% of normal PNP activity exhibit much less severe symptoms of immunodeficiency (Ammann (1978)). However, experience of the clinical toxicity (Siaw et.al. (1980)) associated with the use of the ADA inhibitor 2'-deoxycoformycin ($K_i = 2.5 \times 10^{-12}$ M) (Agarwal et.al. (1977)), which can simulate ADA deficiency in vivo, suggests that PNP inhibitors with K_i values in the 10^{-8} - 10^{-9} range may be preferable as they would allow better modulation of the extent of the inhibition.

CHAPTER FOUR

DATA COLLECTION AND REDUCTION

4.1 Preliminary crystallographic studies

Human erythrocytic PNP is isolated and purified as described in §3.3.1. Crystals are then obtained using the 'hanging drop' method. Droplets consisting of 4 μ l of a 10-20 mg/ml protein solution plus 4 μ l of a 40% saturated solution of ammonium sulphate (A.S.) in 0.05 M citrate buffer, pH 5.3, are suspended on the underside of siliconized coverslips and placed over wells containing 1 ml of a 35-40% saturated A.S. solution in 0.05 M citrate buffer, pH 5.3. After 4 or 5 days at room temperature, large rhombohedral-shaped crystals are obtained (figure 4.1). These crystals have typical dimensions of 0.3-0.4 mm, and the largest grow to 0.6 mm. Rod-like crystals are also obtained at slightly higher pH values. These appear as clusters of elongated rods which appear hexagonal when viewed end on. However, crystals of this form, suitable for X-ray analysis, were never obtained. The rhombohedral crystals can be stored for long periods at room temperature in an artificial mother liquor consisting of 60% saturated A.S. in 0.05 M citrate buffer, pH 5.3.

Experiments by Crate et.al (1980) have established



Figure 4.1 Photograph of the rhombohedral PNP crystals.

that the crystals are enzymatically active. A single crystal was transferred into 170 μ l of a 45% A.S. solution containing starting concentrations of 200 μ M guanine and 17 mM ribose-1-phosphate, and the reaction product guanosine was monitored by following the absorbance at 258 nm, since guanosine has a higher optical density than guanine. The results are shown in figure 4.2. Since the reaction is terminated when the crystal is removed it cannot be attributed to partial dissolution of the crystal. Approximate calculations indicate that the observed activity was much higher than would be expected if only the surface molecules on the crystal were involved. This experiment has been repeated several times with different crystals with the same general results. It is, therefore, reasonable to assume that PNP substrates can diffuse into the crystals and gain access to the active sites on the PNP molecules. This is confirmed by the 6 Å electron density map (5.2) where the active site is seen to have free access to the large solvent channels.

X-ray precession photographs indicate that the crystals are trigonal and are of the Laue group $\bar{3}m$. The space group R32 is specified by the systematic absence of reflections hkl with $-h+k+l \neq 3n$ when indexed according to hexagonal axes. This corresponds to the obverse setting of the rhombohedral cell. The hexagonal unit cell dimensions are $a = 143.8$ Å and

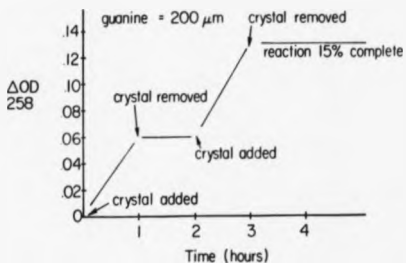


Figure 4.2 Results of Crate et. al. (1980), demonstrating the PNP crystals to be enzymatically active. The optical absorbance at 258 nm follows the concentration of guanosine, the reaction product in this experiment.

$c = 165.1 \text{ \AA}$, with the corresponding primitive rhombohedral unit cell parameters, $a = 99.6 \text{ \AA}$ and $\alpha = 92.4^\circ$. Figure 4.3 shows a 10° precession photograph (resolution limit 4.4 \AA) of the hexagonal $hk10$ zone.

Calculated values of V (Mattheus (1968)) for 1, 2 or 3 subunits per asymmetric unit are 5.47, 2.74, $1.82 \text{ \AA}^3/\text{dalton}$ respectively, indicating solvent fractions by volume of 0.78, 0.56, and 0.33. The density of the crystals was 1.295 g/cm^3 as determined by flotation in a mixture of benzene and ethylene dibromide in 40% A.S. and 0.05 M citrate buffer. The calculated mass per asymmetric unit was then 94000 daltons indicating a complete trimer within that asymmetric unit and a comparatively low (33%) fraction of solvent.

However, when the 6 \AA electron density map was later calculated (see §5), it was immediately apparent that there was, in fact, a single subunit in the asymmetric unit with the three subunits related by the crystallographic 3-fold axis. The accuracy of the experiment to determine the crystal density is the most likely source of the mistake.

Although diffraction spots corresponding to a d spacing of less than 3.0 \AA have been observed on 'still' photographs, the crystals can only be considered to diffract moderately well, and are not particularly stable to X-rays at room temperature. A crystal is



Figure 4.3 Native PNP 10° precession photograph of the hexagonal $hk0$ zone (courtesy of Dr. S.E. Ealick).

effectively 'dead' after approximately 15 hours exposure on a conventional sealed tube or rotating anode generator, allowing at best a single 12° precession photograph, or 1500-1800 low resolution reflections measured on an automatic diffractometer using the method outlined in §5.1. The quality of diffraction beyond about 4.5 Å resolution on such a source is generally very poor and unsuitable for accurate intensity measurement. For this reason it was decided to explore the potential benefits of recording 3-D data using the intense X-ray beam available at the Daresbury SRS.

A report of the preliminary crystallographic study of PNP by Cook et.al. was published in 1981.

4.2 Heavy atom derivative search

The first survey of potential heavy atom derivatives was conducted by the author in the laboratory of Dr. C. E. Bugg. Crystals of PNP were soaked for 1-2 days in 1 ml solutions of artificial mother liquor with the heavy atom reagent dissolved to an initial concentration of 1-10 mM. The effects of the experiments were monitored by screened photography of the rhombohedral 0kl zone (the morphology of the crystals enables the rhombohedral axes to be easily located whereas the hexagonal axes are comparatively difficult to find) using a Supper precession camera mounted on a Rigaku Ru200B rotating anode X-ray generator, operating at

typically 40-45 kV and 30-50 mA.

The iodinated inhibitor 5'-iodoformycin B (IFB) (synthesized by Dr. S-K Chu) was an obvious candidate and did indeed show intensity changes when compared to the native diffraction pattern, despite the relatively low concentration of one iodine atom in 30000 daltons. The only other iodinated substrate or analog available was 8 - iodoguanine (I-G) and this gave the same intensity changes, although smaller in magnitude, seen with the IFB. The diffraction patterns from these two complexes indicated no lack of isomorphism. IFB was later successfully cocrystallized with the enzyme by Dr. W. J. Cook and, again, these crystals were completely isomorphous with the native PNP crystals.

Results of soaks with standard heavy atom reagents were disappointing. Most showed no intensity changes at concentrations just below that which seriously damaged the crystals. The reagents tested are shown in table 4.1. Since there are sulphydryl groups present which are essential for enzymatic activity, it was expected that mercurial derivatives could be prepared. However, of these, only p-chloromercuribenzene sulphonate (PCMBs) showed any potential at all. Generally, PCMBs produced changes in the diffraction pattern only at concentrations which damaged the crystals to some extent. Diffractometer and film data were later collected from this derivative and were successfully used in the 6 Å study. At higher resolution

Table 4.1 Summary of the initial search for suitable heavy atom derivatives using precession photography.

Compound	Soak conditions	Effect on crystals	Effect on diffraction pattern
P-chloromercuri-benzoate PCMB	> 10 μ M < 10 μ M	Cracking No effect	No changes No changes
P-chloromercuri-benzene PCMBB	> 70 μ M 5 μ M 1 μ M 1 day	Cracking within 1 hr Cracked after 1 day No damage	Weak diffraction, changes visible some changes, not reliably reproducible No changes
Mercuric acetate	> 1 μ M < 1 μ M	Immediate cracking No damage	No changes
Mersalyl acid	1 μ M 0.1 μ M 24 hrs	Immediate cracking No damage	No changes
HgCl ₂	1 μ M	Immediate cracking	
K ₂ Pt(CN) ₄	5 mM 1 mM	Immediate cracking Slow cracking in 1 d	V. poor diffraction changes as for PCMBB, but smaller in magnitude
K ₂ PtCl ₆	1 mM 0.1 mM	Cracking after 12 hrs Absorbed yellow/green colour of solution	No changes
K ₂ PtI ₆	10 μ M 1 μ M 1 day	Immediate cracking Slight crazing	No changes
K ₂ Pt(NO ₃) ₆	0.1 mM 20 μ M 1 day	Immediate cracking No damage	No changes
KAu(CN) ₂	>20 μ M 20 μ M 2 days	Immediate cracking No damage	No changes
K ₂ UO ₂ F ₆	5 mM 7 days	Did not absorb yellow of solution	No changes
UO ₂ (Ac) ₂	10 mM 7 days	No damage	No changes
UO ₂ (NO ₃) ₂	50 mM 5 days	No damage	No changes
PbCl ₂	5 mM 3 days	No changes	No changes
PbAc ₂	Saturated 7 d.	No changes	No changes

Table 4.1 (Continued)

Compound	Soak conditions	Effect on crystals	Effect on diffraction pattern
AgNO ₃	10 mM	Slightly crazed	No changes
LuAc ₃	10 mM 3 days	No changes	No changes
DyAc ₃	50 mM 4 days	No changes	No changes
SmAc ₃	5 mM 3 days	No changes	No changes
Sm(NO ₃) ₃	10 mM 4 days	No changes	No changes
SmCl ₃	5 mM 3 days	No changes	No changes
TbCl ₃	50 mM 20 mM 3 days	Cracked in 5 mins Slightly cracked	No changes
NaReO ₄	Saturated 3 d.	No changes	No changes
5'-Iodo formycin B	50% sat 3 days	No changes	Changes observed, but no clear reversal
B-Iodoguanine HCl	50% sat 3 days	No changes	same as IFB

the usefulness of this derivative is somewhat more questionable. Precession photographs of the IFB, 1-G and PCMBs complexes are shown in figure 4.4

The lack of success with the lanthanide reagents was probably due to the citrate buffer in the artificial mother liquor chelating the lanthanide cations. Cacodylate was substituted as the buffer and native crystals were found to be stable at pH 5.2. The diffraction pattern from a crystal so stored for a week, appeared identical to the normal native pattern. However, experiments with SmCl_3 and NaReO_4 failed to show any positive results.

A second phase of heavy atom screening was later conducted by Drs. S. E. Ealick and Y.S. Babu. In this case potential derivatives were tested by collecting a complete 3-D 6 Å data set on an automatic diffractometer, using the methods described briefly in §5.1. By these means a 6 Å data set could be collected from a single crystal in about 18 hours. Similar exposure on the precession camera yields a relatively poor quality 8° or 10° precession photograph. Although this method leads to a large amount of time being wasted in processing the data from unsuitable derivatives, it does avoid the often inconclusive results obtained using precession photography. Use of this method is obviously highly dependent on the necessary diffractometer time being freely available.

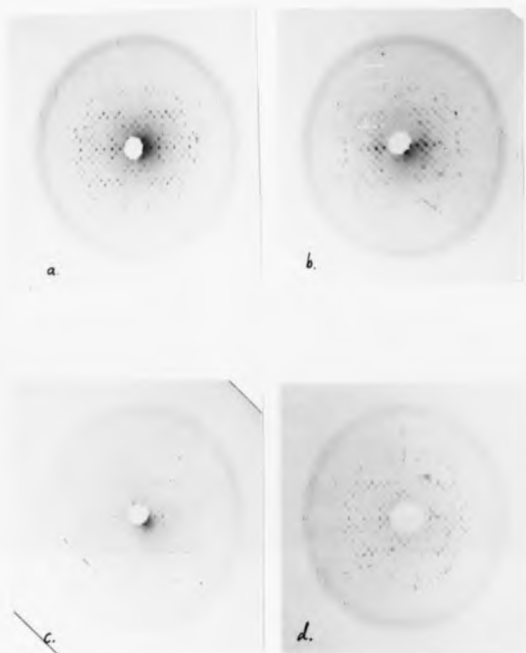


Figure 4.4 Native and derivative PNP 8 0kl precession photographs;

- a) Native, 15 hrs exposure, 40 kV, 45 mA ;
- b) PCMBs 20 M, 24 hr soak, 18 hrs exposure, 40 kV, 40 mA ;
- c) IFB 500 M, 3 day soak, 18 hrs exposure, 45 kV, 30 mA ;
- d) I-G 50% saturated, 3 day soak, 12 hrs exposure, 40 kV, 40 mA.

This search revealed a high quality ethylmercuri-thiosalicylate (EMTS) derivative and two lesser quality derivatives in gold thiomalate (GTM) and gold dicyanide (GCN). Note that the latter reagent was dismissed as unsuitable in the earlier screening thereby, to some extent, justifying the use of the diffractometer methods discussed in the previous paragraph. GTM ($\text{Na}_2\text{AuC}_4\text{H}_3\text{O}_4\text{S}\cdot\text{H}_2\text{O}$) is relatively uncommon as a heavy atom reagent in protein crystallography. It is widely available as an antirheumatic, and the sample used in the work was purchased in a local drug store.

4.3 Data collection at the SRS

Following the decision to use synchrotron radiation for high resolution data collection, a native data set was successfully recorded at the SRS in November, 1981, prior to the installation of the vertically focusing mirror. Four derivatives and a second native data set were recorded in October and November 1982 and the EMTS derivative data followed in February 1984. In addition, another ten data sets (mainly PNP-inhibitor complexes) have recently (June, 1984) been recorded using radiation from the wiggler magnet at wavelengths around 1.0 \AA .

At the time of this work, photographic film was the only detector available for recording protein crystal data at the SRS. Consequently screenless oscillation photography (Arndt et.al. (1973), Arndt

and Wonacott (1977)) was chosen as the most efficient method of collecting the diffraction data. The principle of the method is very simple: the only motion is the rotation of the crystal about an axis through the crystal centre and perpendicular to the X-ray beam. In this way, each reciprocal lattice point is rotated in turn to cut the sphere of reflection and is recorded on a flat stationary film (figure 4.5).

The data are recorded on a series of successive small angle rotation photographs, the angle being chosen to minimize the number of overlapping reflections whilst maximizing the number of fully recorded reflections. Spots which are partially recorded on a film because they occur close to the start or end of the rotation range, will be continued on the adjacent film. Providing the rotation drive mechanism is free of backlash, the two portions may be added together to give an accurate measurement of the reflection intensity. This facility of the method becomes crucially important when large unit cells are studied and the range of rotation per photograph is limited to 0.5 or 0.75°. There is a small but significant blind region along the rotation axis, and this is overcome by either recording equivalent reflections elsewhere in reciprocal space, or by rotating the crystal about a different axis. The total rotation angle needed depends on the space group symmetry - the higher the symmetry the smaller the range needed for the same number of equivalents.

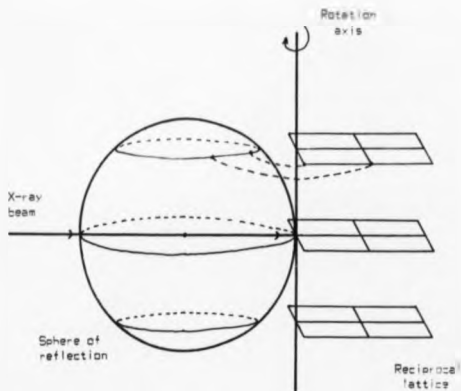


Figure 4.3 Rotation camera geometry. Reciprocal lattice points are brought into the reflecting position by rotation of the crystal about an axis, \perp , perpendicular to the X-ray beam.

For the PNP crystals mounted with a rhombohedral reciprocal cell axis along the capillary (and therefore along the rotation axis), there are no symmetry elements along the rotation axis and so the necessary range is difficult to envisage. 55° of rotation from -5° (relative to a perfectly set crystal) enables a full data set with, on average, 3½-fold multiplicity. However, the multiplicity is widely distributed with, typically, 600 reflections recorded once only, and 400 recorded 8 times or more, after all rejections, in a data set of 11400 unique reflections within the 3.16 Å resolution sphere.

The maximum rotation range per photograph is determined by the separation of adjacent reciprocal lattice points which cause diffraction to the same point of the film (figure 4.6). So, assuming a 3.0 Å resolution limit,

$$\begin{aligned}\theta &= d^*/d_m^* \\ &= d_m/d \\ &= 3.0/99.2 \\ &= 1.73^\circ\end{aligned}$$

This should be reduced by an amount equal to the total reflecting range of the crystal (i.e. the sum of the crystal sample mosaic spread and, in the case of a horizontal rotation axis, the X-ray beam vertical crossfire) and so a value of 1.5° was selected as a

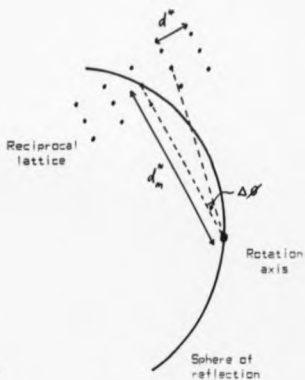


Figure 4.6 The maximum rotation range, $\Delta\beta$, before reflections which overlap on the film are stimulated, is determined by the reciprocal distance corresponding to the required maximum resolution, (d_n^*) , and the appropriate reciprocal lattice spacing (d^*).

good compromise.

The camera that was mounted on the alignment carriage of the instrument was a standard Enraf-Nonius Arndt-Wonacott oscillation camera (Arndt et.al. (1973)) in which the control unit has been modified to provide repeated oscillations for the short exposure times required with SR. The crystal to film distance was close to 95 mm throughout, so that the edge of the film corresponds to a resolution limit of approximately 2.7 Å. The film used was CEA Reflex 25 which combines a low fog level with reasonable speed. To enhance the dynamic range of the recorded data, each film pack consisted of a cassette loaded with three films enabling the weak high angle reflections to be measured on the front (a) film and the strong reflections on the back (c) film.

Before each period of allotted beam time, the internal camera alignment was checked to ensure normal beam geometry. the wavelength was calibrated to 0.001 Å by locating the K absorption edge ($\lambda = 1.488$ Å) in a Ni foil. Regular checks were made to ensure maximum X-ray flux through the collimator (0.6 mm diameter was used throughout) as measured by a simple ionization chamber between the collimator and the sample. The Ge (111) monochromator crystal (oblique cut $\alpha = 10.44^\circ$) was set at the Guinier position where, for this wavelength, $p' = 2.48$ m, thereby ensuring that

the curvature component of the spectral spread was zero. The spectral resolution under these conditions is 1.6×10^{-3} as described in §2.1.2.

The PNP crystals were mounted in clean siliconized glass capillaries sealed at either end, first with wax and then with glue. Plugs of artificial mother liquor on either side of the crystal prevented it from drying out. Crystals were mounted either immediately prior to data collection, or before being transported from the U.S.A. In the latter case there were no problems with crystal deterioration.

The procedure for collecting data from a crystal was as follows:

a) The crystal setting is obtained by taking two still photographs at $\theta = 0^\circ$ and 88° . Arc and spindle corrections are then applied according to the positions on the films of the major circles of spots. If the spindle readings, corresponding to the two 'set' positions, differ by 92° rather than 88° then the spindle setting of 92° is absolute 0° or 180° . The ambiguity is resolved if the layer lines can be seen clearly enough on the stills, or failing this, inversion of the diffraction pattern is easily recognized when processing the data.

b) Orientation photographs are taken at $\theta = 0^\circ$ and 88° . Since Friedel pairs do not, in general, occur on the same film, it is not necessary to ensure that

the crystal is perfectly set, although it does help in overcoming problems when processing the data.

c) The exposure rate (seconds/°) is set according to the beam conditions and automatic data collection commenced. After a full carousel of 8 film packs have been exposed, the first and last two are developed and an estimate made of how many more packs may be exposed before the radiation damage to the crystal becomes excessive. Typically, another 6 packs are exposed with the last two being duplicated by the next crystal.

d) The orientation stills at $\theta = 0^\circ$ and 88° are repeated so that the final crystal orientation is available if any problems with crystal slippage are later detected.

Table 4.2 gives the details relevant to the data sets recorded on the standard bending magnet beam line. No absorption corrections have been applied to the film data collected to date. Indeed, until recently, no one had addressed the problem of measuring crystal absorption surfaces with SR. Helliwell et.al. (1984) describe a method which uses a small ion chamber between sample and collimator to monitor the decaying incident beam. A second, larger, gas filled ion chamber is then used to measure the beam intensity transmitted by the crystal as a function of the spindle axis angle and the inclination angle.

Table 4.2 Details of native and derivative data sets collected at the SRS on station 7.2 .

Crystal Soak	No.	Phi range deg	Exposure sec/deg	Beam Cav	mA	Flux nA	Date	Collected by
Nv 1	1	n/a	n/a	1.8	n/a	n/a	Sept '81	CEB,JRH
	4	6.0-15.0	750	1.8	100	0.305	Nov '81	SRE,TJC
	6	13.5-25.5	750	1.8	65	0.180	"	&SAR,JRH
	7	24.0-36.0	500/750	1.8	120	0.315	"	"
	8	34.5-46.5	500/750	1.8	130	0.350	"	"
	9	355.0-359.5	600	1.8	90	0.265	"	"
IPB 500 μ M 3 day	1	355.5-7.5	50	2.0	110	1.9	Oct '82	SAR,TJC
	2	314.0-332.0	25	2.0	215	4.2	"	&JRH
	3	329.0-347.0	25	2.0	160	3.2	"	"
	4	344.0-2.0	25	2.0	140	2.6	"	"
Nv 2	1	355.0-19.0	25	2.0	150	3.8	Nov '82	SAR,SRE
	2	16.0-40.0	25	2.0	140	3.6	"	&TJC,JRH
	3	28.0-52.0	15	2.0	210	5.1	"	"
	4	11.5-16.0	25	2.0	130	3.2	"	"
		& 43.0-56.5						
IPB CC	1	355.0-19.0	20	2.0	160	4.0	"	"
	2	13.0-37.0	15/20	2.0	180	4.6	"	"
	3	31.0-55.0	20/25	2.0	170	4.5	"	"
	4	7.0-16.0	25	2.0	190	3.7	"	"
		& 28.0-32.5 & 38.5-43.0						
I-G 50% 3 day	1	355.0-19.0	25	2.0	140	3.5	"	"
	2	16.0-34.0	15	2.0	190	4.9	"	"
	3	22.0-46.0	20	2.0	160	4.0	"	"
	4	43.0-52.0	20	2.0	165	4.1	"	"
		& 355.0-7.0						
PCMB5 20 μ M 3 day	1	355.0-19.0	25	2.0	150	3.7	"	"
	2	16.0-37.0	25	2.0	140	3.5	"	"
	3	16.0-34.0	25	2.0	125	3.1	"	"
	4	32.5-50.5	40	2.0	115	2.9	"	"
EMTS 25 μ M 3 day	1	355.0-13.0	200	1.8	62	0.54	Feb '87	SAR,JRH
	2	8.5-20.5	200	1.8	49	0.45	"	& JB
	3	20.5-32.5	300	1.8	40	0.38	"	"
	4	32.5-50.5	300	1.8	32	0.32	"	"

Notes

- Exposures of sub-50 seconds were effected by a single sweep through the θ range, those over 50 seconds consisted of 10 sweeps.
- Flux refers to the output signal from the air-filled ion chamber placed after the collimator.
- The initials in this and the next table (4.3) refer to the following co-workers: C.B. Bugg, S.E. Halick, T.J. Greenhough, J.R. Bellivell & J. Babash.

Fortunately the PNP crystals are block-like rather than plate-like, but nevertheless, variable absorption will introduce errors into the integrated intensities. The data collected on the wiggler station at short wavelengths will perhaps provide some indication as to how serious this problem is.

Typical examples of diffraction patterns taken at the start and towards the end of a crystal lifetime are shown in figure 4.7, along with a pattern recorded on a conventional X-ray source from a fresh good quality crystal. Although it cannot be said that the resolution limit of diffraction by the PNP crystals is improved by using SR, there is no doubt that the quality of the data in the important annulus of $4.5\text{--}3.0\text{ \AA}$ is greatly enhanced. In addition, the volume of data collected per crystal is increased approximately ten-fold, since the crystal providing the conventional source photograph of figure 4.7c did not yield a second useable photograph.

4.4 Data reduction

All the PNP data collected at the SRS have been processed using the MOSCO suite of programs written by Dr. A. J. Wonacott and others (Arndt and Wonacott (1977)). The first native data set was processed at Imperial College and, since that time, the suite has been implemented at Daresbury, where there are currently facilities to process films both on- and

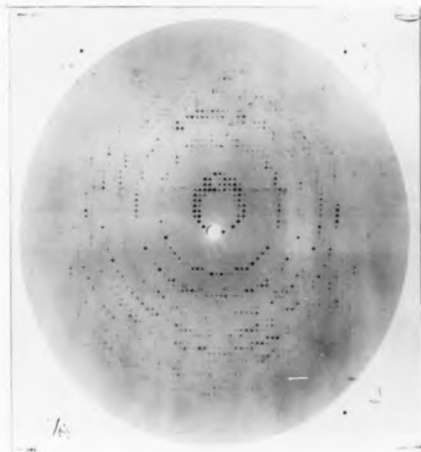


Figure 4.7a Photograph of the first oscillation exposure taken from crystal 1 of the Nv2 data set. Three pairs of 15 sec setting stills had already been taken. The exposure time was 25 sec/deg and the edge of the film corresponds to 2.75 Å resolution.

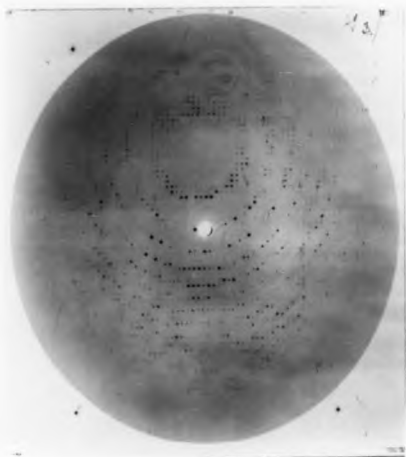


Figure 4.7b The twelfth oscillation photograph recorded from crystal 1 of the Nv 2 data set. Exposure time was again 25 sec/deg.

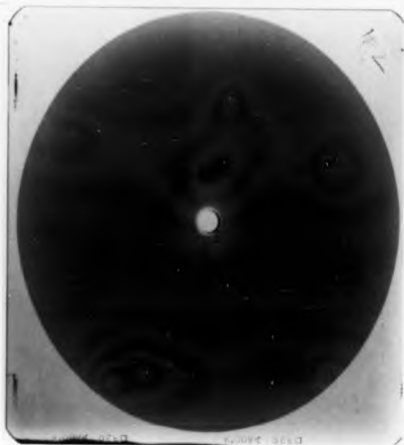


Figure 4.7c A 1.5 degree oscillation photograph taken from a fresh crystal using a rotating anode source operating at 40 kV, 40 mA. The total exposure time was 4 hr 10 min. The edge of the film corresponds to 2.7 Å resolution. A second exposure from this crystal showed only diffuse scatter. (X-ray facilities courtesy of Dr. L. Sawyer)

off-line. The films are digitized by a Joyce-Loebl Scandig 3 microdensitometer controlled by a Data General Nova 3-12 computer. The film image is stored on disk for immediate processing or is written to magnetic tape so that it can be processed at a later date using a VAX 750 computer. The PNP films were all digitized using a 50 μ m raster.

The MOSCO suite consists of the following programs:

a) IDXREF - Given the coordinates of spots from a set of still photographs, indexes the reflections and then refines the crystal orientation, cell parameters, and crystal to film distance.

b) GENVEE - Given the refined parameters output by IDXREF, creates a GENERATE file containing the indices, coordinates, and degree of partiality of all reflections expected to occur within the given rotation range of the crystal.

c) MOSFLM - Firstly, minimizes the discrepancies between the predicted spot positions and the observed centres of gravity of a subset of the observed reflections on the stored film image, then presents an average spot profile around which the user constructs a suitable integration box surrounded by a background region of comparable area, and finally, using the convolution of the predicted pattern with the integration box as a mask, sums the optical density in each spot, subtracts the background density, and writes the integrated

intensities to the GENERATE file.

d) POSTCHK - Takes the partially recorded reflections common to a contiguous pair of processed films, and compares the observed and predicted degree of partiality to assess the crystal orientation. It then creates a file of such reflections, along with their observed partiality, as input to IDXREF for post-refinement.

e) ROTCOR - Corrects the integrated intensities for non-linear film response to intensity, absorption (assuming appropriate data is available) and oblique incidence. Also applies Lorentz and polarization corrections (Kahn et.al. (1983) have developed the expressions for the latter correction that are applicable for SR).

f) PASCAL - Calculates and applies scale factors between films within a film pack. Outputs agreement and intensity statistics.

The first step in processing film data is to obtain refined values for the crystal orientation matrix and missetting angles. Since the Daresbury implementation of MOSCO was not commissioned at the time, it was decided to conduct this first phase in the home laboratory where any problems could be overcome at leisure. To this end the first segment of the OSCAR suite of processing programs was implemented to run in batch mode on the Keele GEC 4082 computer.

To avoid the tedious and error prone task of reading

spot coordinates from the still photographs using a Stoe measuring device, a method based on a Ferranti Cetec digitiser was developed. Films are taped down on the digitiser and the cross wires on the curaoor are positioned over a spot. On pressing the cursor, the x and y coordinates of the crosswire position are written to a disk file. Once the fiducials and a suitable number of evenly distributed spots (excluding the blind region) have been measured, a Fortran routine transforms the coordinates into the appropriate format for input to the OSCAR seg 1 program. The resolution of the digitizer is 0.1 mm - a significant improvement over the Stoe device.

Crystal cell parameters, missetting angles, crystal to film distance and estimates of η , the sample mosaic spread, were all obtained from OSCAR using this method (not without various difficulties, such as missing fiducials and films slipping in the cassette after the fiducial marks were made) and used as the starting parameters in IDXREF. One advantage of IDXREF is that it allows constrained refinement of the cell parameters. In using OSCAR with the trigonal PNP crystals the cell parameters were allowed to refine independent of one another until a satisfactory solution had been obtained and then, a final refinement was run using fixed averaged cell parameters.

The post-refinement facility of MOSCO is an extremely useful method of monitoring the accuracy of the crystal

orientation being used, and will reveal crystal slippage. In processing the PNP data POSTCHK was run between every pair of contiguous film packs. The changes in missetting angles were generally very small and so were allowed to accumulate until they reached some value in the region of $0.02-0.03^\circ$, at which point the remaining film packs were regenerated, using the updated angles. The partials bordering this change were discarded.

To obtain good agreement between observed and predicted partiality, one needs an accurate value for the crystal rocking width, which is the sum of the sample mosaic spread, η , and the X-ray beam crossfire angle, γ . In an attempt to improve the estimate of η , IDXREF was run repeatedly with different values of η , using the reflections output by POSTCHK after the first two film packs of a crystal. The two parameters monitored were the r.m.s. angular deviation (Resid) of reciprocal lattice points from the observed position on the Ewald sphere and the sum of $((R_o - R_c/d^*)^2)$, which is the function minimized in the orientation refinement (R_o and R_c are the observed and calculated distances from the Ewald sphere centre O). In the cases where this was done, these parameters decreased in magnitude as η was reduced, until they levelled off to steady values at $\eta = 0.10 - 0.15^\circ$. In some cases, a shallow minima was observed as in figure 4.8. These results, whilst being far from conclusive, suggest that

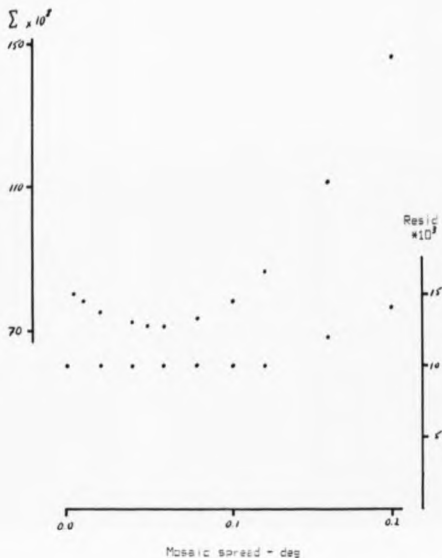


Figure 4.8 Estimation of the sample mosaic spread using IDXREF and the first two film packs of Nv2 crystal 1. Definitions of Resid and \sum can be found in the text. 28 reliable reflections, close to half recorded on each film, were used in the refinement of two mis-setting angles.

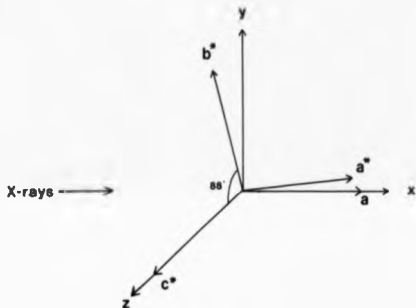
the sample mosaic spread was somewhat smaller than had been thought (generally $0.20 - 0.25^\circ$ was used for other crystals).

Table 4.3 shows the parameters relevant to the starting orientation of each crystal, and indicates which crystals were treated in the above manner. Whilst over-prediction will cause some fully recorded reflections to be flagged as partials, thereby adding noise to the integrated intensities, this is far less serious a problem than under-prediction, where some partials are flagged as fully recorded, and the remaining portion on the next film is ignored. Throughout the data processing, the plots generated by MOSFLM of predicted and observed patterns were monitored very closely for the occurrence of unpredicted reflections, and this was especially so when a low value of η was used. In addition, it is likely that the mosaic spread will tend to increase as the crystal becomes disordered through radiation damage.

At the time of processing the first two datasets (Nv1 and IFB soaked), the MOSCO programs had not been updated to include the changes due to the SR beam geometry. In these cases, the values used for the beam divergence and the spectral spread were 0.2° and 2.5×10^{-3} , respectively. The current versions of the software separate the horizontal and vertical divergence (0.201° and 0.013° respectively) as well as the

Table 4.3 (over page) Crystal orientation data.

The crystal mis-setting angles $\Delta\theta_{x,y,z}$ are relative to a standard crystal setting. In the present case, this is defined such that, at a θ angle of zero, the reciprocal c^* axis is parallel to the rotation axis (z), and the real crystallographic a axis is parallel to the X-ray beam (x), as shown below.



Crystal		θ deg	θ deg	$\Delta\theta$ deg	$\Delta\theta$ deg	$\Delta\theta$ deg	CF mm	θ_{ref}	Resid deg	η deg	By whom
Nv1	1	0.01500	87.778	-0.217	-0.672	-0.543	23.87	-	-	0.25	TJC
	4	0.01502	87.777	0.012	0.292	0.117	94.02	-	-	0.25	-
	6	0.01504	87.764	0.372	-1.528	-0.375	94.29	-	-	0.25	SAR
	7	0.01502	87.843	0.108	0.352	-0.682	94.63	-	-	0.25	-
	8	0.01503	87.765	0.353	-0.092	-0.474	94.60	-	-	0.25	-
	9	0.01503	87.690	0.148	-1.340	-0.065	94.69	-	-	0.25	-
Nv2	1	0.01496	87.724	0.126	-0.963	0.208	96.96	150	0.059	0.12	SAP
	2	0.01496	87.622	-0.293	0.476	-0.168	96.91	150	0.044	0.12	-
	3	0.01495	87.681	0.921	0.258	-0.837	96.89	150	0.051	0.12	-
	4	0.01498	87.515	-0.432	-0.269	-0.169	96.94	150	0.051	0.12	-
IFB	1	0.01501	87.993	-1.963	0.678	-0.629	96.81	95	0.043	0.25	SAP
	2	0.01504	87.899	0.860	-0.184	1.530	97.29	91	0.044	0.20	-
	3	0.01500	87.900	0.441	-0.173	-0.230	97.34	94	0.069	0.20	-
	4	0.01496	87.903	-0.221	0.495	0.579	97.62	96	0.062	0.20	-
IFB CC	1	0.01491	88.061	0.212	-0.291	-0.318	97.06	128	0.032	0.20	TJC
	2	0.01492	87.998	-0.069	-0.111	0.511	97.09	82	0.032	0.20	-
	3	0.01495	87.934	-0.067	-0.128	-0.114	97.00	104	0.036	0.20	-
	4	0.01492	88.002	0.124	0.268	-0.031	96.71	65	0.071	0.15	-
1-G	1	0.01493	87.715	1.644	0.469	-1.124	96.37	150	0.040	0.25	JH
	2	0.01491	87.713	0.483	-0.190	-0.296	96.12	120	0.039	0.25	-
	3	0.01492	87.803	-1.449	-0.712	0.099	96.37	119	0.043	0.25	-
	4	0.01491	87.731	0.427	0.180	-0.048	96.44	131	0.069	0.25	-
PCMS	1	0.01495	87.734	0.678	0.036	0.211	96.27	150	0.044	0.20	JH
	2	0.01496	87.739	0.187	1.004	-0.035	96.36	121	0.054	0.20	-
	3	0.01495	87.733	0.663	0.034	-0.259	97.12	106	0.058	0.20	-
	4	0.01496	87.687	0.169	-0.515	0.565	96.16	115	0.047	0.20	-
EMTB	1	0.01499	87.638	1.067	-1.297	0.131	95.79	91	0.033	0.20	JH
	2	0.01498	87.637	-0.047	-0.465	-0.097	95.67	76	0.050	0.20	-
	3	0.01494	87.651	-0.173	-0.402	0.320	95.59	89	0.044	0.20	-
	4	0.01498	87.633	-0.824	-0.224	-0.596	95.54	71	0.035	0.20	-

Notes

1. Nv1 was processed at Imperial College before the introduction of POSTCHK and the incorporation of synchrotron beam geometry.
2. IFB was processed before the incorporation of synchrotron beam geometry.
3. CF is the refined crystal-to-film distance.
4. θ_{ref} is the number of reflections used for refinement in IODREF.
5. Resid is defined in the text.
6. η is the sample mosaic spread used in the data processing. A value of 0.15 or less invariably indicates that the effect of varying η was investigated.

conventional (1.6×10^{-3}) and correlated (zero) components of the spectral dispersion, effectively making use of the understanding of the beam geometry gained in developing the basis for the polychromatic profile method.

Table 4.4 shows the results of the merging of film packs as output by the program PASCAL. The most noticeable feature of these statistics is the wide variation in film factors (F_{AB} , F_{AC}) suggesting inconsistency in the film itself or in the developing conditions. The developing solutions were replaced regularly and the recommended development times were adhered to, leaving variations in the temperature of the solutions, and the lack of suitable agitation in the developer both as the most likely causes of variation in the developing conditions. Agreement factors (R) between intensities measured on the three films within a pack are normally considered satisfactory at 4 - 5%, and in general the PNP data shows values of 4 - 6%. The EMTS crystals are a notable exception for which there is no apparent explanation.

The data from the individual film packs were merged together and reduced to a unique set of reflections using the programs ROTAVATA and AGROVATA, which are implemented as part of the CCP4 protein crystallography program suite at Daresbury. ROTAVATA calculates scale and temperature factors between overlapping batches of data, by the method of Fox and Holmes (1966). AGROVATA applies the scale factors, adds together partially

Table 4.4 Results of merging films within a pack averaged over all packs from each crystal

Crystal	N packs	Scales A/B	A/C	R factor %	$\Sigma I/\Sigma I_0$	Mean I	Mean I ₀	N sym pairs	R sym %
Nv1	1 4	3.15(08)	—	9.1	40.7(6.3)	598(41)	1708	102	12.3
	4 6	2.66(01)	—	5.4	44.9(6.3)	794(66)	2129	48	9.1
	6 8	2.56(12)	—	4.3	40.2(7.3)	659(42)	2137	65	16.6
	7 8	2.50(04)	—	5.0	47.7(3.7)	668(50)	1624	122	18.9
	8 6	2.65(06)	—	4.7	46.5(4.2)	642(39)	1625	147	6.9
	9 3	2.68(03)	—	4.2	47.6(1.9)	672(24)	1662	23	11.5
Nv2	1 16	2.56(12)	6.06(44)	4.1	52.2(4.4)	916(65)	1590	646	7.8
	2 8	2.47(23)	5.85(59)	4.3	50.6(4.4)	858(65)	1536	175	6.8
	3 15	2.72(42)	7.10(11.50)	5.3	41.7(4.8)	822(85)	1594	502	13.3
	* 11	2.35(21)	5.64(97)	5.3	41.1(4.7)	681(55)	1526	494	14.4
IFB	1 8	2.40(05)	5.89(34)	7.8	50.5(6.9)	1049(87)	1820	198	9.6
	2 10	2.95(20)	8.71(64)	5.4	43.4(5.4)	843(45)	1812	155	10.6
	3 12	2.86(08)	8.07(38)	4.5	51.8(2.8)	797(94)	1825	44	9.5
	4 12	2.70(08)	7.45(19)	4.2	44.1(5.4)	656(58)	1823	265	15.3
IFB	1 13	2.55(10)	6.34(42)	5.4	41.7(3.1)	705(59)	1598	548	14.1
	CC 2 13	2.48(16)	6.12(40)	5.2	43.4(4.1)	811(49)	1544	159	14.2
	3 16	2.72(32)	6.69(32)	6.1	46.2(3.6)	648(50)	1590	115	11.4
	4 16	2.45(08)	5.85(22)	5.1	47.2(3.2)	849(64)	1541	276	12.2
I-C	1 16	2.53(09)	6.23(29)	4.7	53.3(5.6)	1011(74)	1580	557	8.3
	2 13	2.55(10)	6.37(52)	5.2	32.0(12.3)	712(144)	1627	153	28.3
	3 16	2.51(11)	5.95(50)	5.1	36.3(4.5)	744(108)	1644	391	17.8
	* 12	2.53(09)	6.11(55)	5.0	36.9(5.2)	708(93)	1631	714	22.2
PCMBS	1 16	2.64(17)	5.70(47)	5.9	37.4(6.0)	788(90)	1611	322	13.8
	2 12	2.59(13)	6.67(47)	5.1	34.5(6.8)	730(66)	1627	379	17.4
	3 12	2.47(13)	6.37(1.02)	5.6	40.4(6.4)	873(68)	1635	149	18.6
	4 12	2.50(10)	5.90(29)	4.9	43.0(5.3)	914(106)	1628	331	11.8
EMTS	1 9	2.44(05)	5.95(21)	8.0	45.4(8.6)	877(107)	1600	327	10.4
	2 8	2.39(10)	5.66(22)	9.8	36.1(4.2)	697(46)	1613	82	23.2
	3 6	2.40(08)	5.60(24)	9.3	32.3(5.4)	705(64)	1605	83	15.5
	4 8	2.42(11)	5.98(26)	8.4	38.4(4.7)	825(52)	1621	227	18.9

Notes

1. Nsym and Psym are the number and R factor of symmetry pairs occurring on the same film.

2. R factor is defined as

$$\sum_{hkl} \frac{|I - \bar{I}|}{\sum I}$$

where \bar{I} is the mean intensity of the N reflections with intensities I , and common indices h, k, l .

recorded reflections, rejects bad agreements between repeated measurements or symmetry equivalents, averages the repeated data, and finally, constructs a file of unique hkl's with structure factor amplitudes and anomalous differences. In this process any film packs showing particularly poor agreement with the remainder of the data set were discarded and these were invariably the last data to be collected from the crystal in question. Since radiation damage manifests itself as a fall off in intensity with increasing resolution, the temperature factors calculated by the scaling program provide an indication of the damage sustained by a crystal. Figure 4.9 plots the temperature factors assigned to the film packs for each crystal in the Nv2 dataset. The trend is approximately linear and of similar rate for each sample. Crystals in the other datasets gave very similar results.

Initially, the two native datasets were merged independently, but after careful analysis, the data were combined into a single dataset at the film pack stage. The mean fractional difference in the structure factors between these two data sets is constant at 6.7% up to 4 Å resolution, above which it steadily increases. This observation is common to all data sets and is discussed in §4.5. In addition, difference Fourier maps, calculated using $F_{Nv1} - F_{Nv2}$ coefficients, and SIR phases calculated using the IFB derivative, revealed

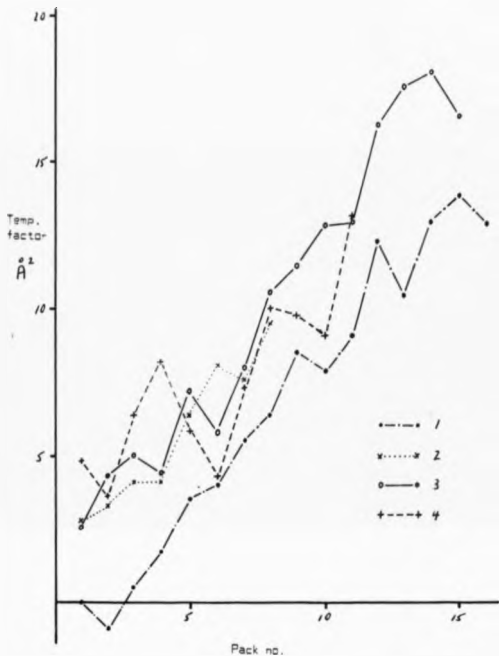


Figure 4.9 Temperature factors for each film pack from the four crystals used in the Nv2 data set, plotted as a function of pack number.

no significant features apart from the ghost peak at the iodine site. The results of the merging of film packs for all data sets are given in table 4.5.

The agreement (R) factors between equivalent measurements appear rather high, 5-10% generally being regarded as acceptable. However, as figure 4.10 demonstrates, for the native data set, the inherently weak (and therefore less well determined) high angle data is responsible for inflating the overall R factor. The same trend is demonstrated for the IFB cocrystallized data in table 4.6. Another cause for concern is the systematic tendency for the intensities from summed partial recordings to be larger than those from fulls. This has been observed previously for both conventional source and SR source data (Irwin et.al. (1976), Schmid et.al. (1981), Wilson et.al. (1983)). The most likely explanations for this are that:

- a) Errors in the crystal orientation lead to underprediction, causing some reflections, flagged as fulls, to be underestimated as described above, and
- b) Diffuse scatter, having a greater reflecting range than a Bragg reflection, will tend to be spread across two films so, in the case of partials, would be fully measured whereas, for fully recorded reflections, a reduced volume is measured. However, with one exception, this partial bias is not excessively large. No explanation can be given for the large bias apparent

Table 4.5 Intensity statistics for the merging of film packs within each data set.

	Nv 1	Nv 2	Nv comb	IFB	IFB CC	I-G	PCMB	EMTS
Discard	3	8	15	5	9	24	14	6
Nin	49196	53669	96668	56787	66661	44722	51529	27182
Nf	18727	35843	52083	22041	41786	25569	29320	16368
Np	11331	7149	16680	14211	9557	7935	9071	4065
Nout	10100	11424	11424	10495	11414	11341	11370	10208
σf %	14.9	16.0	15.9	17.2	17.5	16.9	18.9	16.7
σp %	18.5	18.5	18.2	23.9	19.6	19.5	20.8	18.8
σ all %	16.4	16.4	16.5	20.0	17.9	17.5	19.4	17.2
R %	11.1	11.7	11.6	12.0	12.2	11.7	13.1	11.9
Bias %	-2.3	-2.6	-2.0	-9.0	-1.8	-3.1	-2.2	0.9
$N > 3\sigma$ %	73.0	74.4	78.9	77.0	74.4	71.7	69.8	69.1
$N\Delta$	-	-	-	7890	8153	7961	8442	4797
$\Delta > 3\sigma$ %	-	-	-	4.1	3.7	1.4	1.6	5.7

Notes

1. All data outside the 3.16 Å resolution shell were discarded on input.
 2. Discard is the number of packs discarded due to poor agreement with the remainder within the data set.
 3. N refers to numbers of reflections: f, p refer to fulls and partial pairs respectively.
 4. σ refers to standard deviations, and Δ to anomalous differences.
 5. R is defined as in table 4.4 .
 6. Bias is defined as $\sum |I_i - \bar{I}| / \sum I_i$
 7. A reflection was tested for rejections if any measurement of that reflection differed from the weighted mean by more than 3σ , in which case the following criteria were applied.
 - If N is the number of measurements of a reflection, then :
 - if $N=2$, reject the smaller;
 - if $N=3$, reject that which lies on the opposite side of the unweighted mean to the others;
 - if $N \geq 4$, try the $N=3$ test and if no rejection is made then reject that for which $|I_{\text{mean}} - I|/\sigma$ is greatest.
- If the reflection again fails the 3σ test repeat the rejection process unless more 25% of the measurements of the reflection have already been rejected.

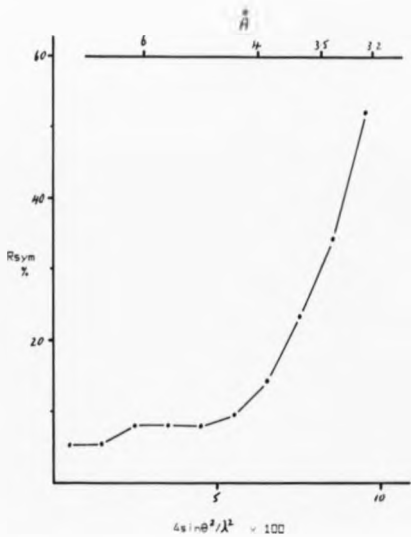


Figure 4.10 a The R_{sym} of the combined native data as a function of resolution.

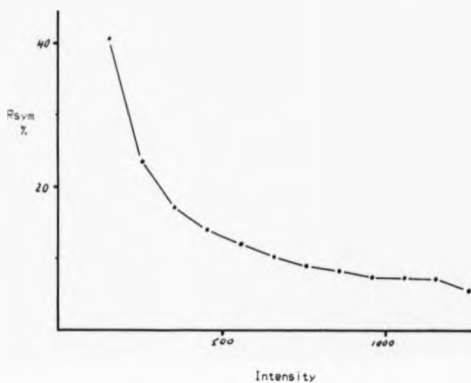


Figure 4.10 b The R_{sym} of the combined native data as a function of intensity.

Resolution	Nout	$\Sigma 3\sigma$	σ/I	R
$\frac{\pi}{A}$		Σ	Σ	Σ
3.16	11414	74.4	17.9	12.2
3.20	10874	76.3	16.8	11.6
3.30	9917	79.2	15.6	10.7
3.40	9105	81.7	14.5	9.9
3.50	8356	83.6	13.6	9.3

Table 4.6 Variation of merging statistics with resolution limit for the IFB CC data set.

Data set	Scale	Temperature Factor k^2
IFB	1.086 (32)	9.3 (2.4)
IFB CC	1.051 (21)	6.0 (1.7)
I-G	1.052 (09)	5.3 (0.8)
PCMB5	1.039 (09)	4.0 (0.7)
EMTS	1.053 (11)	5.2 (1.3)

Table 4.7 Residual relative scale and temperature factors for each derivative (with standard deviations) after local scaling to the native structure amplitudes

in the soaked data set - a large value of η was used in processing the data and the films show no comparative excess of diffuse scatter.

The final step in the processing of the PNP data was to convert the rhombohedral indices to those of the hexagonal cell (see International Tables for X-ray Crystallography Vol. 1 §1.5).

A simple Fortran routine was written to make the conversion, transpose the reflections into a unique region of the hexagonal reciprocal lattice, and to sort the reflections. Care was taken to ensure that the consistent sense of the anomalous data was preserved.

4.5 Analysis of the derivatives

The various derivative data sets were scaled to the native data by making the ratio $I|F_p|/I|F_{PH}| = 1.0$, using the 3D option in the program ANSC. Reflections were grouped together, by level along the l axis, and in ranges of ρa^2 (where ρa is the cylindrical radius) and of ϕ (measured from the h axis). The ranges were adjusted to ensure each block of reciprocal space contained at least 50 reflections, and then the above ratio was analysed for each block. The scale factors were carefully monitored as a function of each index, of intensity, and of $S (= 4 \sin^2 \theta / \lambda^2)$, both before and after scaling to ensure no systematic variations, and no such problems were encountered for any of the derivatives.

The residual scales and effective temperature factors are listed in table 4.7. Figure 4.11 shows the variation of k_{emp} as a function of S , and figure 4.12 shows the mean fractional isomorphous change in F , for all the derivatives. The latter plots clearly demonstrate the progressive increase in the fractional changes at resolutions beyond 4.0 \AA , indicating that the errors in the data are substantial at high angle, there being no indication of any lack of isomorphism. At low angle the changes in the I-G and the PCMBs data are hardly larger than those seen between native data sets, suggesting that these are not good quality derivatives (see §6 for further discussion). However, the results for the IFB and the ENTS indicate good heavy atom substitution.

The IFB soaked derivative was the first to be studied at a time when the other data sets were yet to be reduced. An isomorphous difference Patterson synthesis to a resolution of 6 \AA was calculated when the data was still indexed according to the rhombohedral cell. A solution with up to three sites in the asymmetric unit was anticipated since a complete trimer was presumed to occupy that region. In fact, a single site solution - which accounted for nearly all the major peaks in the Patterson map - was obtained (figure 4.13). The site refined moderately well using the centric terms only and this enabled a rough 6 \AA phase set to be calculated. From this point the remaining

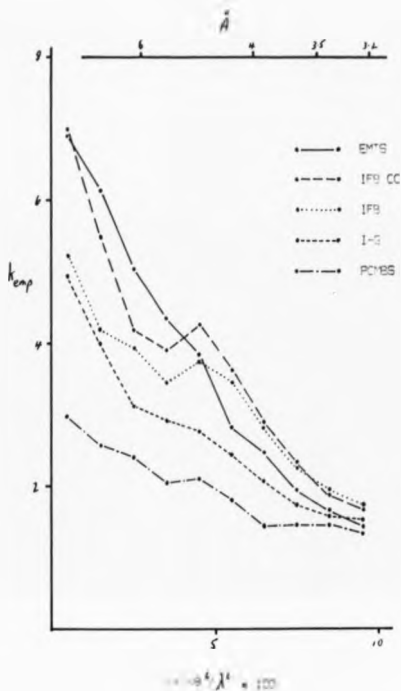


Figure 4.11 The variation of k_{sep} as a function of S for each derivative. The low values at high angle suggest that the anomalous differences are greatly overestimated.

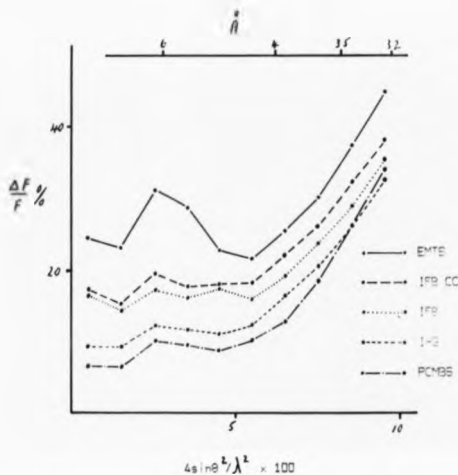
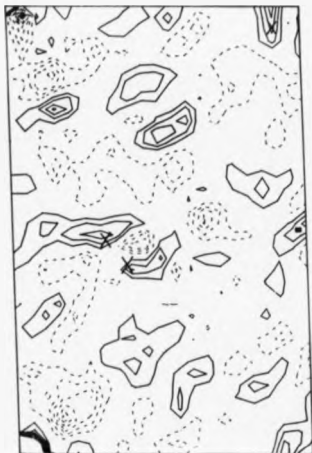


Figure 4.12 The mean fractional isomorphous difference plotted as a function of S for each of the derivative data sets.

$$\frac{\Delta F}{F} = \frac{\sum (|F_{m1}| - |F_p|)}{\sum |F_p|}$$

0,0,1



0,0,0

1/2,1/2,0

Figure 4.13 The u - v section from the IFB difference Patterson map calculated in the rhombohedral space group. For each heavy atom in the asymmetric unit vectors occur within this section at $(u=x, y, v=2z)$, $(u=y, x, v=2x)$, and $(u=x, z, v=2y)$. The crosses mark the vectors associated with the subsequently refined co-ordinates of the iodine atom.

derivatives were solved by cross difference Fourier maps and cross-checking with the isomorphous difference Patterson maps. Further details of the heavy atom structures are given in the following chapters.

4.6 Data collection on the wiggler beam line

At the time of this data collection, the development of the instrument on the wiggler beam line was still at an early stage. The vertically focusing mirror had not been installed, thereby limiting the X-ray flux available. The Enraf-Nonius FAST TV detector was being commissioned on a sealed tube source, and so the Arndt-Nonacott oscillation camera used previously was mounted on the alignment carriage on the wiggler instrument.

It was decided to use a wavelength of 1.02 \AA in order to a) maximize the X-ray flux (see §1.3.1), b) reduce sample absorption, and c) optimize the anomalous dispersion effects in the gold derivative. GTM. (Au L_{III} edge, $\lambda = 1.0403 \text{ \AA}$). A Si (111) monochromator, with an oblique cut of 6.75° was set at the Guinier position for which $p' = 3.45 \text{ m}$. The monochromator vessel was filled with helium at atmospheric pressure to increase the dissipation of heat in the monochromator environment. The wavelength was calibrated by scanning the Au L_{III} absorption edge in a $4 \text{ }\mu\text{m}$ gold foil.

A total of 22 eight hour shifts over a period of ten days, were used to collect all the data sets listed

in table 4.8. Since the SRS was operating 24 hours a day, two teams of three people (SEE, TJG, YSB and SAR, JRH, JH) worked alternately to maintain efficiency over the entire period. Nine data sets were recorded before the wavelength was moved to 0.99 \AA to collect an EMTS data set (Hg L_{III} edge, $\lambda = 1.00944 \text{ \AA}$). The final data set was recorded from a single 8 - AGua crystal where the exposure time per film pack was reduced to less than half of what would otherwise be used, and a full 32 films packs were exposed. This was done to try and establish whether the advantage of not having to merge the data from a number of different crystals outweighs the disadvantage of under-exposing each film pack.

At the time of these experiments, the SRS was operating at 2 GeV with injection currents of 200-250 mA. Exposure times were between 150 and 300 seconds per degree of crystal rotation, depending on the stored current at the time. All data sets (save the last) required 3 or 4 crystals and there was little evidence of any improvement in crystal lifetime at these wavelengths. However, until the above data have been processed and fully analysed no further conclusions can be drawn.

Data set	λ -Å	Function
Native	1.02	Reference
Gold thiomalate	"	Derivative
Gold cyanide	"	"
Guanine	"	Substrate
8-aminoguanine	"	Substrate analogue
8-iodoguanine	"	"
5'-iodo-9-deazainosine	"	"
Formycin B	"	"
Acyclovir	"	"
EMTS	0.99	Derivative
8-aminoguanine	"	Rapid data collection

Table 4.8 Data sets recorded on film in June 1984 using the protein crystallography workstation on the wiggler beamline at the SRS.

CHAPTER FIVE

THE 6 Å STRUCTURE OF PNP

The processing of the 3 Å high resolution data, collected at the SRS, was conducted concurrently with the second stage of heavy atom screening (§4.2). This latter work led to the calculation of a high quality 6 Å electron density map, as well as spotlighting the importance of the EMTS derivative, from which 3 Å resolution data were collected as the 6 Å map was interpreted.

5.1 Data collection and calculation of phases

6 Å datasets (each from a single crystal) were measured by Drs. S. E. Ealick and Y. S. Babu using a Picker FACS-1 automatic diffractometer. The method used was based on that suggested by Hansen et.al. (1979). Measurements were made by step-scanning through each reflection and then applying a profile fitting technique to obtain integrated intensities. Reflections were precalculated and sorted into shells of 100 reflections such that the total angle driving time on the diffractometer was minimized. The highest resolution shell was collected first and the lowest resolution shell last, to minimize the resolution dependent effects of crystal decay. The details of the step-scans and the profile fitting are given by Ealick et.al. (1984). Standard reflections were monitored after each shell to provide the means for applying a decomposition correction, and absorption corrections were also applied

using the method of North et.al. (1968). Multiple independent datasets were measured from native crystals and most of the derivatives to improve the quality of the data. In addition to the heavy atom derivatives described here, diffractometer data were collected from a number of crystals soaked in solutions containing PNP substrates or substrate analogs (see §5.3), as well as several heavy atom reagents that failed to produce derivatives.

Once reduced to a set of structure amplitudes, each derivative data set was scaled to the native data by applying an overall scale factor and an isotropic temperature factor. As described in §4.5, the interpretation of the IFB isomorphous Patterson map by the author, using the SRS film data, enabled all the heavy atom sites in the remaining derivatives to be located in cross difference Fourier maps. These sites were then refined by least squares techniques using data from the centric zones (approximately 250 reflections to 6\AA). Phases for the Fourier calculations were obtained using independent derivatives to ensure no confusion between true peaks and ghost peaks in the maps. Table 5-1 lists the derivatives used in the final phase calculation. The PCMB5-C derivative is the 6\AA subset of the SRS data set and the IFB data sets (soaked and recrystallized) consist of diffractometer and SRS film data averaged together. Table 5-2 compares the refinement parameter for the active site complexes. The refinement statistics of the IFB derivative are significantly improved

Derivative	Conditions	No. crystals	$\Delta F\%$
IPB	50 % Sat. or co-crystallised	3	16.3
PCIBS 'A' + Guanosine	2.5 μ M + 1 mM	1	28.6
PCIBS 'B'	25 μ M	2	28.2
PCIBS 'C'	20 μ M	4	7.9
EDTS	0.1 mM	1	28.4
GTH	0.6 mM	2	12.3
GCN	0.1 mM	2	9.7

Table 5.1 Heavy atom derivatives for which 6 Å diffractometer data were collected and used to calculate the 6 Å protein phases. All soaks were for 3 days prior to data collection. The mean fractional change, ΔF , was defined in figure 4.12 .

Complex	Model	Relative Occupancy	Fractional Co-ordinates			Rc %	F_H/E
			x	y	z		
IFB	I	10.55	0.104	0.183	0.135	67	1.04
IFB	I	6.95	0.114	0.197	0.135	57	1.39
	S	8.60	0.097	0.162	0.133		
	P	9.88	0.086	0.140	0.159		
IFB CC	I	8.69	0.116	0.197	0.133	55	1.42
	S	10.90	0.093	0.156	0.139		
	P	7.56	0.087	0.141	0.162		
IFB + IFB CC	I	7.64	0.115	0.197	0.136	55	1.42
	S	9.00	0.097	0.162	0.134		
	P	9.36	0.086	0.138	0.159		
I-Guanine	I	8.04	0.084	0.136	0.166	60	1.24
Guanine	I	10.52	0.083	0.130	0.166	60	1.13

Table 5.2 Results of the refinement using centric reflections to 6 Å resolution, of the substrate- and substrate analogue-PNP binary complexes. F_H/E is the average value of the calculated heavy atom structure factor amplitude divided by the average lack of closure error. Rc is the residual, $\sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$, where F_{obs} and F_{calc} refer to the observed and calculated heavy atom structure factors. For the IFB analysis the sugar (S) and the base (B) were each modelled by a single iodine atom and fixed isotropic temperature factors (15 Å^2) were used throughout.

when the IFB molecule is modelled by three atoms corresponding to the centres of the observed of the purine, the sugar, and the iodine moiety. This model was the result of the studies described below (§5.3). In addition, there appears to be no significant differences between data collected from samples soaked in IFB solutions and from those cocrystallized with IFB, and the data were accordingly averaged together. The refined heavy atom parameters for the derivatives used in the phase calculation are shown in table 5.3. All the heavy atom positions were assigned fixed isotropic temperature factors of 6.0 \AA^2 . It can be seen that mercurial and gold atoms bind in varying combinations and relative occupancies to three sites within the crystallographic asymmetric unit. It is also clear that there is only one accessible active site within the asymmetric unit; moreover, the lack of further sites in the co-crystallized PNP + IFB complex suggests that there is a single monomer within the asymmetric unit. The table also suggests that the IFB and the EMTS are the better derivatives whilst the PCMBs and the GCN are of a poorer quality. The correct enantiomorph of the structure was determined using the anomalous signal in the SRS film data, as is described in §6.2. Protein MIR phases were calculated by the method of Blow and Crick (1959) using the program PHASE (§6.3 describes the use of this program). The mean figure of merit for the resulting phase set was 0.794. Finally, an electron density map was calculated using the native structure amplitudes

Derivative	Site	Relative Occupancy	Fractional Co-ordinates			Rc %	F _H /E
			x	y	z		
PCMBs 'A'	a	10.76	0.194	0.193	0.119	62	1.08
	b	7.19	0.186	0.037	0.072		
	c	4.71	0.005	0.082	0.239		
PCMBs 'B'	a	14.10	0.193	0.202	0.121	60	1.16
	b	3.79	0.181	0.049	0.082		
	c	1.58	0.027	0.107	0.230		
PCMBs 'C'	a	3.61	0.192	0.201	0.121	67	1.06
BNTS	a	13.14	0.218	0.196	0.134	47	1.96
	c	7.66	0.028	0.123	0.197		
GTH	b	6.22	0.161	0.030	0.088	58	1.24
GCN	a	3.29	0.204	0.202	0.125	63	0.89

Table 5.3 Results of the refinement of the non-active site heavy atom derivatives. The notes for table 5.2 apply here also. The protein phases were calculated using the combined IFB data set in addition to these six derivative data sets.

with the MIR phases and figure of merit weighting.

5.2 The electron density map at 6.0 Å resolution

Examination of the electron density map indicates that the PNP molecule is a trimer with the identical subunits related by the crystallographic three-fold axis, thereby disproving the earlier conclusion (54-1) that the asymmetric unit contains a complete trimer. Figure 5-1 is a projection down the three-fold axis from $z = 0.0$ to 0.26 (fractional cell coordinates). The largest peak in the map is of height 0.21 (arbitrary units) and figure 5-1 is contoured at 0.05 and 0.10 . The molecule is very clearly resolved from the solvent regions suggesting that the map is of good quality. The unit cell contains an unusually large amount of solvent. Based on the calculated molecular weight of 32300 daltons and a partial-specific-volume for the protein of $0.73 \text{ cm}^3/\text{g}$, the unit cell is 78% solvent by volume.

Pairs of trimers form face to face hexamers across a crystallographic two-fold axis. Other two-fold axes relate subunits from adjacent trimers and serve to delineate the molecular boundary as in figure 5-1. Figure 5-2a shows a schematic representation of the unit cell packing, viewed down the three-fold axis, in which each subunit is represented by a sphere of 40 Å diameter. A large solvent cavity, approximately 110 Å in diameter and 80 Å thick, is formed by a circle of six hexameric units and is capped top and bottom by additional hexamers. Figure 5-2b shows

Figure 5.1 Projection of the PNP 6 Å electron density map viewed parallel to the trimer axis. The dimensions of the projected density are approximately 215 x 124 x 45 Å thick. The three lines represent the crystallographic diads which occur at $z=1/6$.

Figure 5.2 Stereoview of the subunit packing in the PNP crystals. Each subunit is represented by a sphere of approximately 40 Å diameter, and the crystallographic unit cell is represented by a box; a) view parallel to the trimer axis, b) view down the major solvent channel which runs continuously through the crystal lattice.

Figure 5.3 Projection of a central slice of the 6 Å electron density, viewed parallel to the trimer axis. The dimensions are 150 Å, along the crystallographic 'a' axis, x 110 Å x approximately 15 Å thick. The overlay indicates four rods of density identified as possible α -helices packed around a β sheet.

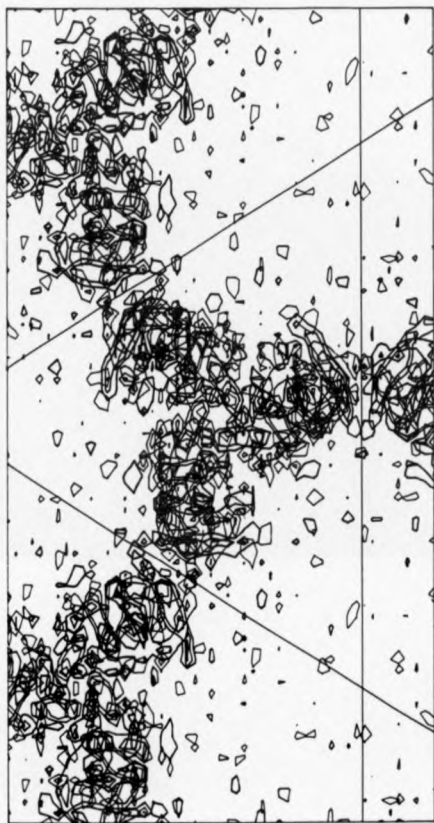


Figure 5.1

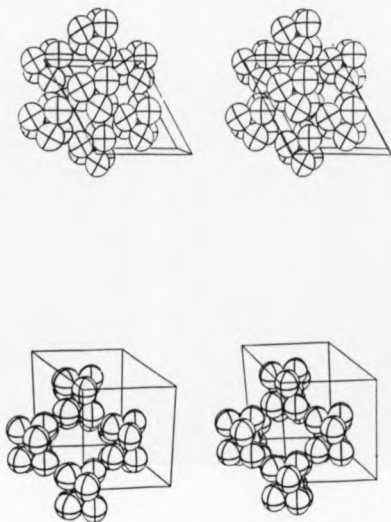


Figure 5.2 a) upper b) lower



Figure 5.3

a similar representation in which the view is down the major solvent channel. This channel has a minimum cross-sectional dimension of about $50 \text{ \AA} \times 80 \text{ \AA}$ and connects the larger solvent cavities together in a continuous series.

As a result of this packing arrangement, the substrate binding site, as well as most of the molecular surface, is accessible to the solvent. This is consistent with the finding that the crystals have enzymatic activity (0.4-1). The heavy atom binding sites are near the surface of the molecule well away from any subunit-subunit boundaries. Human erythrocytic PNP has four sulfhydryls per subunit (see §3.3.4) and it is likely that the metal positions correspond to three of these. The major PCMBS and EMTS site is located near the active site and studies have shown that the first sulfhydryl group titrated by PCMB causes an 80% inactivation of the enzyme. Conversely GTM binds away from the active site and recent studies have shown that GTM does not fully inactivate the enzyme even at 10^{-3} M (Stoeckler et al (1984)).

Although positioning of individual amino acid residues is not possible at this resolution, a number of secondary structural features are recognisable. There are several columns of electron density of a shape and diameter appropriate to that of an α -helix. Figure 5.3 shows a projection down the three-fold axis from $z = 0.07$ to 0.16 (i.e. about 15 \AA thick). The overlay indicates the electron density associated with four of the potential helices. These

are packed in such a way as to suggest the region of electron density between them is perhaps part of a region of β sheet and that this central region of the subunit resembles the organisation of the parallel α/β configuration seen in many known protein structures, e.g. Lactate dehydrogenase, Flavodoxin and Arabinose-binding protein (Richardson (1981)).

All of these possible helices have been displayed on an Evans and Sutherland Multipicture System using the molecular graphics program FRODO and compared to α -helical polyalanine models. One such example is shown in figure 5.4, and the others have a similar appearance.

5.3 Substrate binding studies

In order to study the PNP-substrate complexes of interest a MIR phase set was calculated using all the derivatives in table 5.1 with the exception of the IFB. Hence, no active site derivative was used and disturbances at the active site were minimized. These phases were used to calculate difference Fourier maps where the coefficients were based on differences between the structure amplitudes of, say, the PNP-IFB complex and those of native PNP.

The difference map computed with the IFB differences to 6.0 Å shows a single extended peak, with the remainder of the map showing no major features at all. There are no large regions of negative density to indicate conformational changes in the enzyme. The IFB density has a significantly larger peak height towards one end and this is assumed to

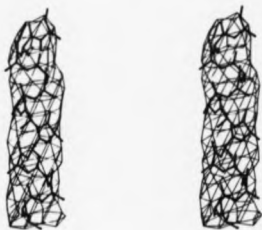


Figure 5.4 Stereoview of one of the rods of density thought to be an α -helix. An ideal poly-alanine α -helical model has been fitted to the electron density to show the dimensions.

correspond to the iodine position. The density was displayed on the Multipicture system and, starting from the iodine atom, an IFB model was fitted to the density. This is shown in figure 5-5. The density has the correct general size and shape for a nucleoside molecule in an anti-conformation (where the glycosidic torsion angle, $O - C(1) - C(9) - N(8)$, lies between 0° and $+90^\circ$, in this case it is approximately 40°). However, at this limited resolution it appears that other conformations also give reasonable fits. Figure 5-6 shows a photograph of the native PNP map (plotted onto spaced acetate sheets) with the IFB difference density superimposed in blue. The active site of the enzyme is clearly seen as an exposed cleft in the surface of the molecule and is possibly formed by the interface of adjacent subunits. It appears that the purine end of the nucleoside fits farthest into the active site pocket, with the ribose moiety lying closer to the solvent interface. The $5'$ - iodo position of IFB is most exposed to the solvent. These observations would be consistent with the bi-bi mechanism previously proposed (§3.3.4).

Difference maps for the FB and the guanine differences have also been computed and these are shown in figures 5-7 and 5-8. In these figures, the IFB model has been left in position, and the IFB density has been replaced by the FB and then the guanine density. These figures show the purine moiety of the model to be within the region of electron density common to all three maps, but at this resolution it is impossible to speculate on the possible



Figure 5.5 Two stereoviews of the 6 Å resolution IFB difference density with an IFB model which has been fitted to this density. The lower view is from the solvent region in towards the active site pocket, i.e. approximately along the crystallographic -b direction.



Figure 5.6 Photograph of the 6 Å protein mini-map with the IFB difference density added in blue, thereby indicating the region of the active site.



Figure 5.7 Two stereoviews of the formycin B difference density superimposed on the IFB model which has been left as fitted to the IFB difference density.

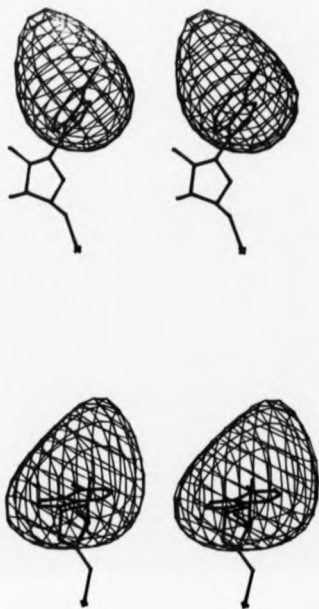


Figure 5.8 Two stereoviews as in figure 5.7 but this time showing the guanine difference density.

differences in the precise positioning and orientation of the substrates.

A difference map, calculated with the data from a crystal treated with both PCMBs and guanosine, failed to show any indication of density in the active site region and this is consistent with the knowledge that sulfhydryl reagents, reacting at heavy atom site A, inactivate the enzyme. It should be possible to treat the enzyme with both GTM and a substrate. This has been attempted with only FB to date, but the difference map showed little density at the gold site. A map calculated from 8 - iodoguanine differences interestingly failed to indicate the presence of an iodine atom in the expected (or any other) position. Investigations are under way to confirm the presence of bound iodine in the material used for the crystallographic experiments. This result also raises further doubts about the I-G derivative data collected at the SRS.

Whilst these low resolution studies are of great interest, more meaningful results can only come from investigations at higher resolution. The data collected on the wiggler beam line (4.6) will constitute the first steps towards achieving that goal.

CHAPTER SIX

THE 3.16 Å STRUCTURE OF PNP

Using the knowledge gained from the studies at lower resolution the heavy atom parameters, protein phases and electron density map were extended in successive stages to 4.0, 3.5 and finally 3.16 Å resolution. This chapter details that work and describes the preliminary interpretation of the protein maps. Since the EMTS data collected at the SRS were not available until after these studies were under way, additional diffractometer data were collected from the EMTS and GTM derivatives in the resolution shell of 6.0 - 4.0 Å using the methods described in §5.1 (Ealick, pers. comm.). Diffraction by the PNP crystals on the sealed tube source of the diffractometer in this resolution range is very weak and the errors in the intensity measurements are consequently high, but, since only one measurement per unique reflection was made and standard deviations were not estimated, it is difficult to indicate quantitatively the quality of the data. Nevertheless, the data did prove useful in the calculation of phases. Once the SRS EMTS data set became available all diffractometer data beyond 6 Å resolution were omitted from the phase calculations.

6.1 Heavy atom parameters

Since approximate heavy atom positions were known for all the derivatives from the 6 Å study the first step was to calculate isomorphous difference Patterson maps at 4.0 Å

resolution to check the self consistency of the Harker vectors and obtain more accurate coordinates for the heavy atoms. The Harker sections and Harker vectors for the space group are given in table 6.1. Figure 6.1 shows the unique regions of the u , v , 0 and the u , $-u$, w sections of the difference Patterson map at 4.0 \AA resolution for each of the derivative data sets recorded at the SRS. The vectors associated with the subsequently refined heavy atom positions are marked on each plot by crosses. In the case of IFB, the iodine vectors are those marked. In the u , $-u$, w sections the scale of the u , $-u$ direction has been halved, the true dimensions are 248 \AA (u , $-u$) by 27.5 \AA ($w/6$). The I-G sections clearly fail to show a solution that agrees with the expected site of substitution. The final plot in figure 6.1 is the $w = 6/100$ section of the EMTS difference Patterson map, showing the expected peak at the vector between the two independent sites.

Once SIRAS phases from the IFB derivative were available cross difference Fourier maps were calculated for the other derivatives to obtain further estimates of the heavy atom positions and to search for minor sites of substitution. Figure 6.2 shows a section of the map calculated with the EMTS differences containing the less well occupied EMTS site. In this map the major site peak is 11.1 (arbitrary units), the minor site is 7.6 and the largest peak in the background noise is 2.6 . The PCMBs difference Fourier map clearly shows a single site of substitution in the crystallographic asymmetric unit. As the studies progressed beyond 4 \AA resolution the Patterson and Fourier maps were not repeated; the heavy atom

Equivalent positions in the R32 hexagonal cell	x, y, z $y, x, -z$ $-y, x-y, z$ $-z, y-x, z$ $y-x, -x, z$ $x-y, -y, -z$	Origins at 0,0,0 ; 1/3,2/3,2/3 ; and 2/3,1/3,1/3 .
Equivalents give rise to 30 vectors on these sections in Patterson space w.r.t. each origin	$u, v, 0 \quad (12)$ $u, -u, v \quad (6)$ $u, 2u, v \quad (6)$ $2u, u, v \quad (6)$	Repeats at $w=1/3, 2/3$) Origins appear)) on these)) sections
$u, v, 0$	$x-y, 2y-x, 0$ $x+y, 2x-y, 0$ $2x-y, x+y, 0$ $2y-x, x+y, 0$ $x-2y, 2x-y, 0$ $y-2x, 2y-x, 0$	+ inverses total 12 equivalents
$u, -u, v$	$x-y, y-x, 2z$ $-x, x, 2z$ $y, -y, 2z$	+ inverses & origin shifts, total 6 triplets each of constant w

Table 6.1 Equivalent positions in real space, along with Harker sections and Harker vectors in Patterson space, for the hexagonal R32 unit cell.

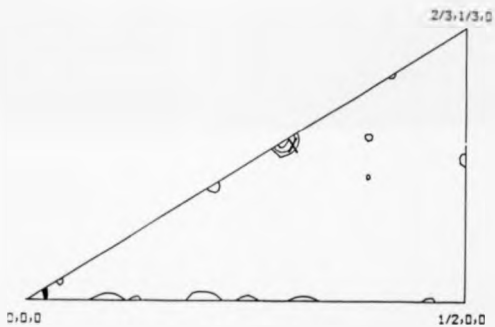
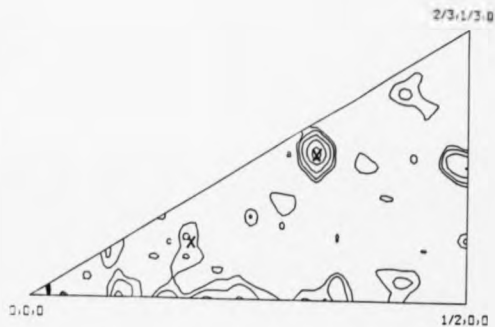


Figure 6.1 $u,v,0$ difference Patterson sections; upper - EHTS,
lower - PCHBS.

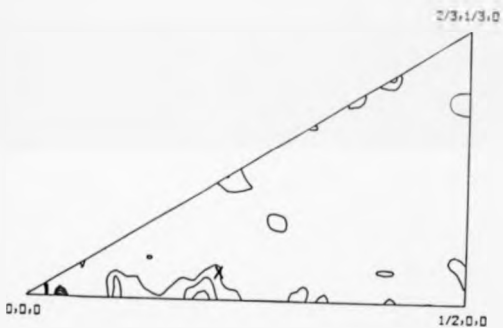
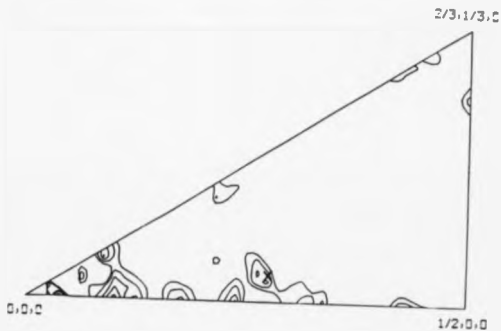


Figure 6.1 (cont) $u,v,0$ difference Patterson sections; upper - IPB,
lower - I-G.

0.1.1.1/6



0.1.0

1.0.0

0.1.1.1/6



0.1.0

1.0.0

Figure 6.1 (cont) u,-u,v difference Patterson sections; upper-BMTs, lower-PCMS

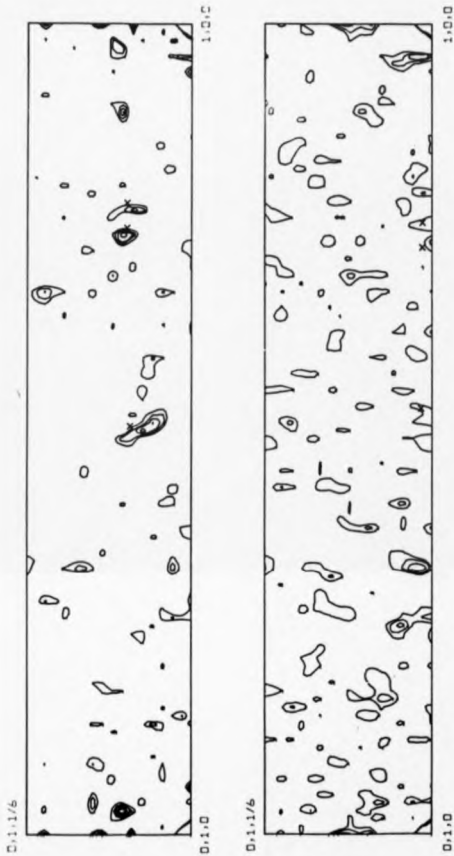


Figure 6.1 (cont) u, -u, v difference Patterson sections; upper-IFB, lower-I-G

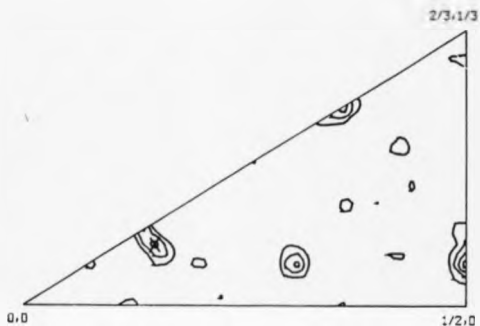


Figure 6.1 (cont) The $v = -6/100$ section from the EMTS difference Patterson map, showing the cross-vector between heavy atom sites 'a' and 'c'.

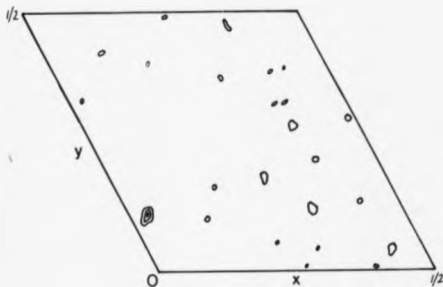


Figure 6.2 The $z = 20/100$ section from the EMTS cross-difference Fourier map, showing the second site (c) of mercurial substitution at (0.03, 0.12, 0.20).

refinements were simply extended to include the higher resolution terms.

All heavy atom refinements used the SCFMLE option in the program REFINE. There are sufficient centric terms to enable utilization of the colinearity of F_p and F_{pH} in a centric least squares refinement. Reflections with $F < \sigma(F)$ were excluded from the refinement as were terms with spuriously large isomorphous differences. Initially fixed isotropic temperature factors were used, these were subsequently allowed to refine, and finally anisotropic temperature factors were introduced. Scale factors calculated by comparison of centric terms alone were used and typical values were around 1.03. No attempt was made to use phased refinement although independent SIRAS phases were available.

Tables 6.2 and 6.3 show the results of the heavy atom refinements, principally at 4.0 and 3.16 Å.

6.2 The absolute configuration

The enantiomorph was initially determined prior to the 6 Å structure determination using the anomalous scattering information in the SRS IFB dataset and comparing the amplitude of the mercury peak in PCMB5 difference Fourier maps. The details below, however, refer to calculations made during the high resolution structure determination, with the same conclusion being reached.

The reversal of the sign of a set of anomalous differences corresponds to a change of enantiomorph since the sign of the measured anomalous difference is absolute. SIR, SIRAS (+),

Derivative	Res. Å	Sites	Nref	Rc %	g	F _n /E	Scale
GTW *	4.0	b	359	62.7	0.22	1.10	1.010
ENTS *	4.0	a,c	491	52.5	0.38	1.63	1.032
ENTS	3.5	a,c	486	46.2	0.62	1.90	1.030
ENTS	3.16	a,c	558	49.2	0.34	1.80	1.030
PCMS	4.0	a	521	62.2	0.21	1.22	1.007
PCMS	3.16	a	780	65.1	0.17	1.14	1.007
IFB	4.0	active	564	62.4	0.36	1.38	1.022
IFB	3.16	active	828	60.3	0.26	1.29	1.022
I-G	3.16	active	706	61.5	0.20	1.20	1.010

Table 6.2 Statistics from the centric refinement of the heavy atom parameters. Diffractometer data sets are denoted by **. Definitions are as before (see table 5.2), with the addition of g, the gradient of the least squares lines between F_{obs} and F_{calc}, and of the derivative/native scale factor.

Deriv.	Res. Å	Rel. Occ.	Fract. Co-ords.			Aniso. Temp. params. $\times 10^3$						
			x	y	z	11	22	33	12	13	23	
GTH *	4.0	10.2	0.164	0.036	0.088	101	66	84	12	37	39	
EMTS *	4.0	13.5	0.217	0.195	0.136	119	111	67	139	-13	-26	
		8.8	0.027	0.120	0.199	-21	160	-13	-17	33	3	
EMTS	3.5	11.5	0.216	0.193	0.132	86	143	105	58	0	-33	
		8.3	0.031	0.121	0.199	132	171	31	-56	-18	29	
EMTS	3.16	9.4	0.215	0.193	0.131	8	35	48	12	-7	-14	
		7.7	0.029	0.121	0.199	87	112	39	-1	0	21	
PCNBS	4.0	3.0	0.190	0.182	0.117	-19	205	53	83	63	-13	
PCNBS	3.16	2.9	0.189	0.182	0.116	20	133	28	64	43	-12	
IFB	4.0	10.1	0.107	0.184	0.134				0 44 0			
		6.7	0.069	0.137	0.161				0 75 0			
		5.5	0.098	0.144	0.140				0 18 0			
IFB	3.16	9.5	0.107	0.184	0.135				0 44 0			
		6.6	0.069	0.135	0.163				0 75 0			
		5.8	0.098	0.143	0.141				0 18 0			
I-G	3.16	5.4	0.077	0.152	0.165	454	224	186	331	-107	-146	

Table 6.3 The final refined heavy atom parameters. The derivatives marked ** are those measured on the diffractometer prior to the availability of the EMTS SRS data set. The three entries for IFB are the iodine, base and sugar moieties respectively. The IFB failed to refine satisfactorily with anisotropic temperature factors: those quoted (0 0) are isotropic values. At 3.16 Å resolution these were held fixed at their 4.0 Å values.

SIRAS(-) phase sets to 4.0 \AA were calculated from the IFB soaked, IFB cocrystallized, and a merged IFB dataset (all used the 3 atom model). The latter data set is a σ -weighted mean of the former two data sets. These phase sets were used to calculate difference Fourier maps with coefficients $F_{\text{EMTS}} - F_{\text{NV}}$. The peak heights of the two EMTS sites are tabulated in table 6.4. The '+' hand is clearly seen to be the correct hand.

The same procedure was attempted using phases calculated from the PCMBs and the I-G data sets. IFB difference coefficients were used with the PCMBs phases and EMTS differences with the I-G phases. These results are also shown in table 6.4. The PCMBs anomalous data did give a marginal indication of the correct hand, although the peak heights are significantly less than the PCMBs ghost peaks. The I-G SIR phases revealed the two EMTS sites but the minor site peak was higher than that for the major site. When the anomalous data were included in the phase calculation on either hand the difference maps failed to distinguish either site from the background noise. This confirms the doubts expressed about this derivative both when analysing the data and in the 6 \AA study. Consequently the I-G data was not used in the subsequent phase calculations.

It was decided that the merged IFB data set would be used in the calculation of protein phases. Table 6.4 suggests that the merged data gives a slightly stronger indication of the enantiomorph. Figure 6.3 plots the ratio of F_{Hcalc} and F_{Hcalc}^* to the mean lack of closure error (E) after SIRAS

Phase set	Href	Anom. hand	FOM	Peak heights		
				Site 'a'	Site 'c'	Iodine
IFB soak	4935	0	37.5	7.30	4.65	(1.87)
		+	46.6	9.15	6.02	(1.81)
		-	45.6	6.78	3.94	(1.93)
IFB OC	5488	0	38.0	7.97	4.61	(1.79)
		+	45.6	9.80	6.28	(1.84)
		-	45.4	7.11	3.57	(2.02)
IFB HER	5523	0	37.6	8.01	4.81	(1.70)
		+	48.5	10.60	6.83	(1.77)
		-	48.4	6.87	3.58	(1.96)
I-G	5391	0	38.1	1.21	2.24	(4.06)
		+	50.1	1.07 *	1.35 *	(4.13)
		-	50.1	1.19 *	0.96 *	(3.88)
PCNBS	5451	0	37.9	(5.81)	-	2.42
		+	43.7	(6.27)	-	2.72
		-	43.7	(6.29)	-	2.38

Table 6.4 Peak heights in 4.0 Å resolution difference Fourier maps calculated using various SIR and SIRAS phase sets. For the IFB and I-G phase sets EMTS differences were used, and for the PCNBS phase sets IFB differences were employed. The peak heights in brackets are those of the 'ghost' peaks, and those marked '*' are at the same level as the local noise.

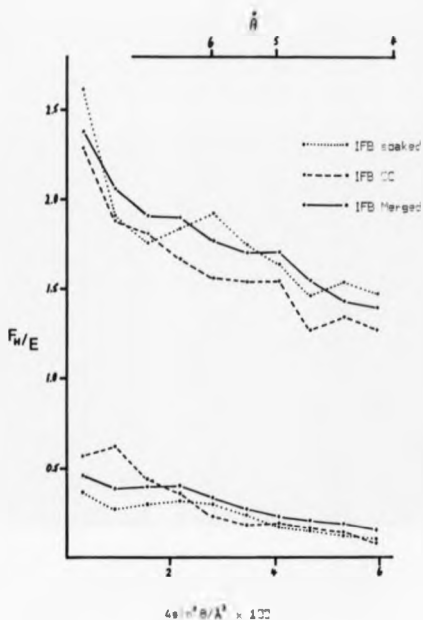


Figure 6.3 The ratio of the mean calculated heavy atom structure factor to the mean lack of closure error in ranges of $4\sin^2\theta/\lambda^2$ for the IFB data sets after SIRAS phase calculation. The upper curves are the isomorphous values and the lower are the anomalous values.

phase calculation on the correct hand for the three data sets and again indicates that the anomalous data at least is slightly improved. At 6 Å resolution there were no significant discrepancies between the difference Fourier maps calculated with the IFB soaked and cocrystallized isomorphous differences, and the protein structure at that resolution gives no reason for thinking the binding of the IFB molecule should be any different since the active site is so accessible in the crystals.

6.3 The calculation of phases

Protein phases were calculated by the statistical method of Blow and Crick (1959) using the program PHASE. These authors showed that a Fourier electron density map computed with the 'best' phase (the centroid of the phase probability distribution) gives the lowest r.m.s. error in the electron density. The phase probability distribution is estimated by minimizing the lack of closure error for each source of information at discrete intervals (10° in this present work). The program requires as input average lack of closure errors in ranges of $4\sin^2\theta/\lambda^2$ along with all the appropriate heavy atom parameters. The anomalous lack of closure errors (E'') were initially set to $E/3$ and a preliminary phase set calculated. The output r.m.s. E and E'' were then input to a second phase calculation and the procedure was repeated until convergence was achieved. In addition the derivative overall scale factors were adjusted to ensure $\alpha_p - \alpha_H$ was on average 90° , i.e. there should be no correlation between

α_p and α_H .

Table 6.5 shows the derivatives used at each stage of phase calculation. Figure 6.4 shows plots of r.m.s. F_{Hcalc} , F_{Hcalc}^* , E , and E^* against $4\sin\theta/\lambda$ for each derivative in the 3.16 Å phase calculation. Figure 6.5 gives an analysis of the figure of merit (m) for the 3.16 Å phases. Figure 6.6 shows the number of reflections in 10° intervals of $\alpha_p - \alpha_H$, from which it appears that the EMTS derivative tends to dominate the phases (α_p is the centroid of the phase probability distribution), whilst the PCMBs derivative contributes least to the final protein phases.

6.4 The electron density map

Electron density maps were calculated using the program FORDAP (written by Ibers and Hamilton) which utilizes a conventional (i.e. slow) Fourier summation algorithm. Figure of merit weights were used and the sampling interval was 1.3 Å in the 4 Å map, 1.15 Å at 3.5 Å and 1.0 Å at 3.16 Å. The maps were plotted in sections perpendicular to the crystallographic c axis onto clear acetate sheets and these were mounted in appropriately spaced cardboard frames. Maps were also plotted in sections perpendicular to the a axis to provide additional perspective when attempting to trace the polypeptide chain. The electron density in these maps is in arbitrary units and was contoured in increments of 2 units starting at 2. The peak values were 19.5, 21.6 and 24.1 for the 4.0, 3.5 and 3.16 Å resolution maps respectively. Subjectively the maps are of high quality. The noise

		Resolution - Å		
		4.0	3.5	3.16
Derivatives included	IFB	IFB	IFB	IFB
	PCNBS	PCNBS	PCNBS	PCNBS
	ENTS *	ENTS	ENTS	ENTS
	GTN *			
No. reflections	5455	8195	10968	
Mean Figure of Merit x 100	68.5	67.6	64.3	

Table 6.5 Summary of the data included at each stage of the phase calculations as the PNP electron density map was extended to 3.16 Å resolution. Diffractometer data is indicated by '*', the remaining data being SRS film data. Anomalous contributions for all the film data were included.

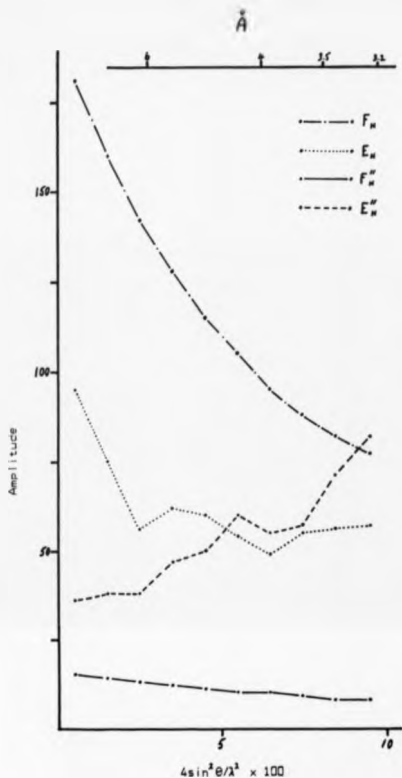


Figure 6.4 Variation with $4\sin^2\theta/\lambda^2$ of the normal and anomalous mean heavy atom scattering factors, along with their associated mean lack of closure errors, for each derivative following MIRAS phase calculation to 3.16 \AA resolution. a) BMTS.

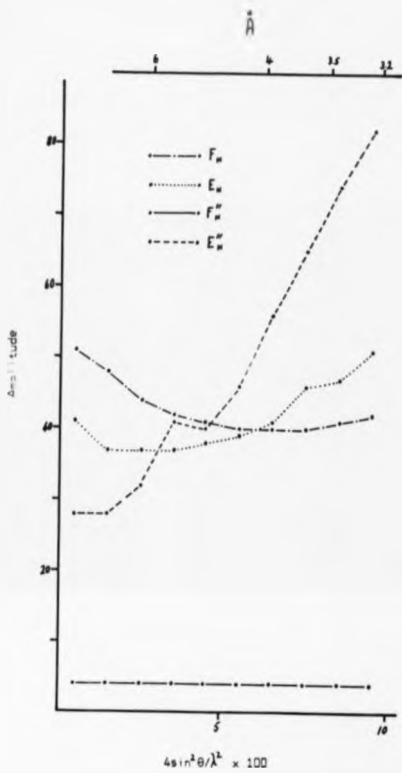


Figure 6.4 (continued) b) PCMBs .

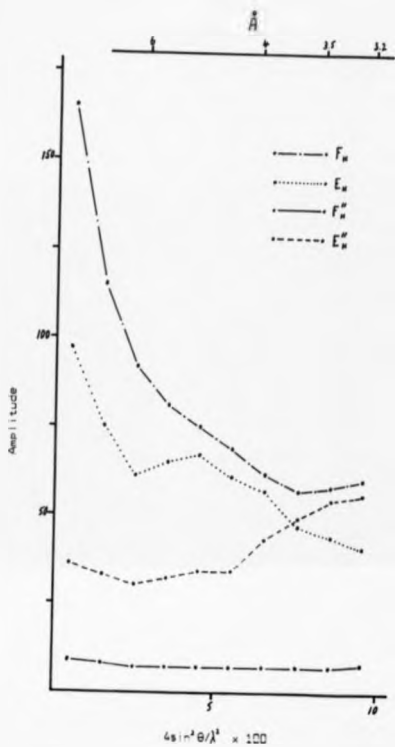


Figure 6.4 (continued) c) IFB

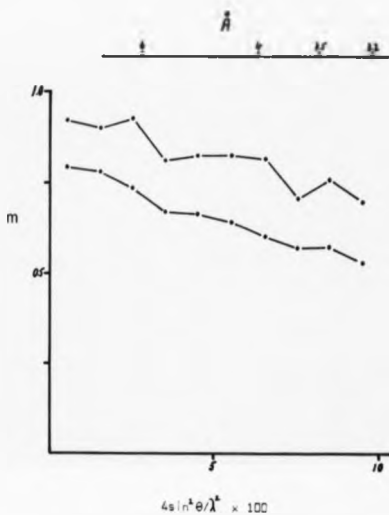


Figure 6.5 a) Variation of the mean figure of merit (m) with $4\sin^2\theta/\lambda^2$ for the centric (upper curve) and acentric reflections as calculated by the method of Blow and Crick (1959).

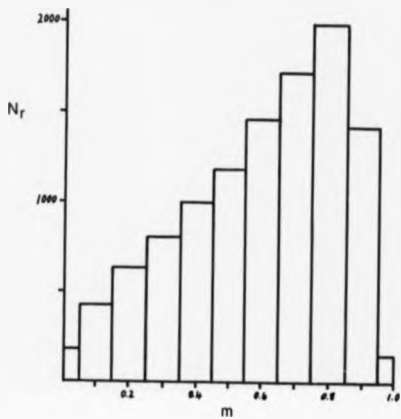


Figure 6.5 b) Figure of merit histogram for the protein phases to 3.16 Å resolution.

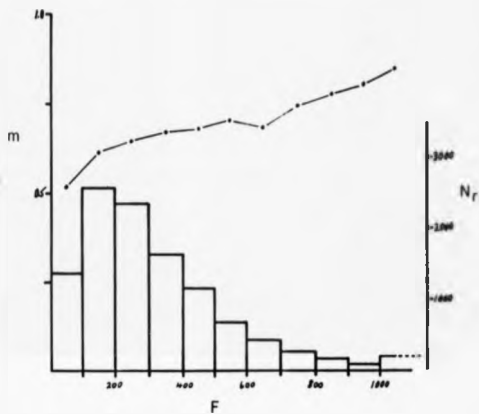


Figure 6.5 c) Figure of merit as a function of native protein structure factor amplitude (F), along with a histogram of the amplitudes

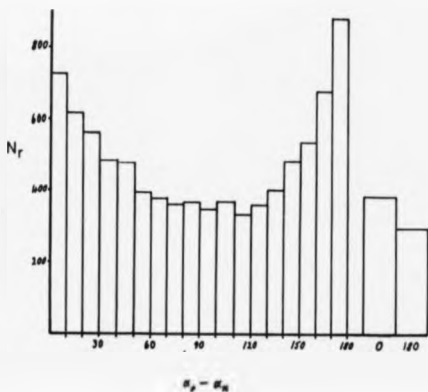


Figure 6.6 Histogram of protein phase angle minus heavy atom phase angle for each derivative in the 3.16 Å phase calculation.

a) EMTS

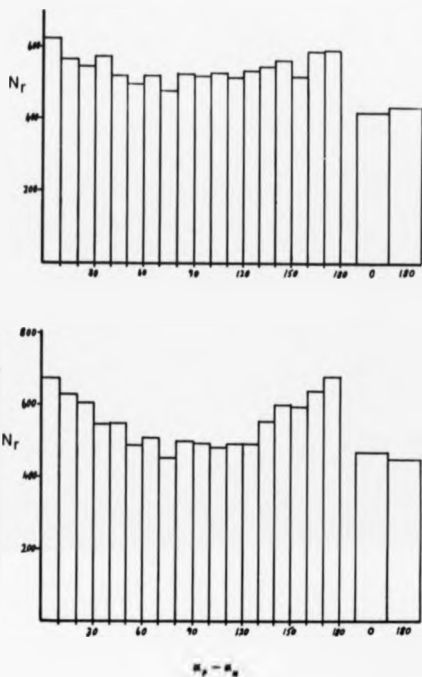


Figure 6.6 (continued)

b) upper - PCMBs

c) lower - IFB

in the solvent regions is of a low level as can be seen in figure 6.7, (taken from the 4.0 Å map) and there is very little density on the symmetry axes. There is a lot of detail in the 4.0 Å map and this increases as the resolution is extended. Large portions of the polypeptide backbone can be traced upon casual inspection. A tentative chain trace through the 4.0 Å map was completed in half a day by placing α -carbon markers along the chain with roughly 3.5-4.0 Å separation, and in doing so the number of markers used corresponded almost exactly to the number of residues in the subunit sequence (288, see appendix I). Inspection of the subsequent higher resolution maps did not indicate any changes in the path of the chain. There were two regions of possible ambiguity in the 4.0 Å resolution map but the later maps did not significantly clarify the proposed connectivity and nor did they enhance the claims of the possible alternatives. In fact the only significant difference between the maps was the improvement in side chain definition at the higher resolution. The regions of density which in the 6 Å map are suggestive of α -helix can clearly be seen to be so in the 4.0 Å map. At 3.5 Å side chains along these helices, by pointing in a consistent direction away from the helix axis, indicate the N to C terminal direction in six of the seven. The seventh, comparatively short, helix is close to the major mercurial binding site and the side chain density is relatively indistinct in this region.

The main bulk of the enzyme subunit consists of a six

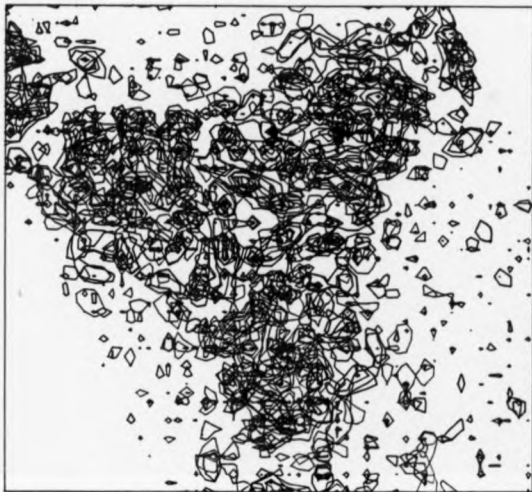


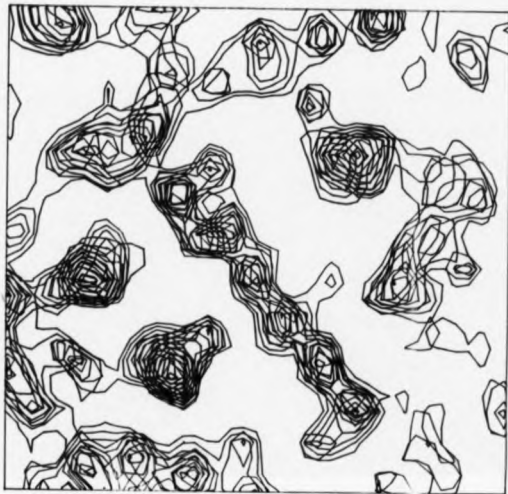
Figure 6.7 A projection, approximately 36 Å thick, down the crystallographic 'c' axis in the 4.0 Å resolution protein map, showing a complete PNP trimer. The map was contoured at 5, 9 and 13 (arbitrary units) and the highest peak was 19.5. This figure shows a similar region of the unit cell to the 6 Å resolution section of figure 5.3.

stranded pleated β sheet displaying the characteristic twist (Caothia, (1973)) against which six helices are packed; two on either side lying approximately parallel to the strands and one lying across both top and bottom of the sheet. The seventh helix (labelled C) packs close to the crystallographic c axis against the three-fold related equivalents. A crude model of the recognizable secondary structure elements was quickly constructed using lengths of dowelling to represent helices and aluminium ribbon for the β strands. Coordinates of the ends of each strand and helix were taken from the 4.0 Å map and brass rods were used to position the components relative to the baseboard, which represents the plane of the crystallographic a and b axes. A photograph of this model is shown in figure 6.8 along with a sketch of the labelling scheme for the secondary structural elements.

The β sheet consists of parallel strands between five and seven residues in length, the exception being strand 2 which is antiparallel if the current chain trace remains unaltered. The parallel strands point towards the active site, that is they run from their amino terminus towards the carboxyl terminus at the active site. Figure 6.9 is a section approximately 4.5 Å thick across the centre of the sheet. The six strands can clearly be seen along with the four helices which pack against the sheet. A portion of helix A can also be seen towards the bottom of the figure. Figure 6.10 shows the adjacent 4.5 Å section on the amino terminus side of the previous figure. Connectivity can be seen between helix E and strand 1, helix C and strand 6, and



Figure 6.8 A photograph of the simple model constructed to represent the secondary structure of the PNP subunit, along with a sketch showing the labelling of the helices (A-G) and β strands (1-6).



100

Figure 6.9 A section from the 4.0 Å resolution protein map, 4.5 Å thick, taken perpendicular to the crystallographic 'a' axis and across the centre of the six stranded sheet, which is clearly seen in the centre of the figure. Four helices, two either side of the sheet, can also be seen in section.

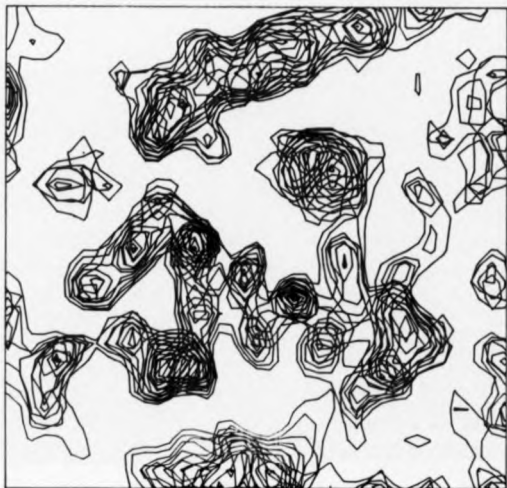


Figure 6.10 The adjacent section of electron density to that of figure 6.9, displaced along the 'a' axis away from the three-fold axis, shows the connectivity between the N-termini of the edge strands of the β sheet and the neighbouring helices.

helix B and strand 5. Helix D can also be seen lying across the top of the sheet. Figure 6.11 shows views (3 Å thick) looking down the c axis, of strands 3 and 5. These sections show the characteristic pleat of the β strands. Figures 6-9 and 6-10 are taken from the 3.5 Å map and figure 6-11 from 3.16 Å map.

The sheet and helices account for approximately half the amino acid residues in the subunit. Some residues form the loops and turns that connect the secondary structural element while the remainder form regions of random coil that are concentrated in the region between helix ζ and the $\alpha\beta$ domain. The active site is a cleft formed between the domain and the region of random coil, and is close to the carboxyl termini of strands 4, 5 and 6.

Although the three dimensional structure can be regarded as consisting of two domains, the amino acid sequence does not divide quite so neatly. In basic terms the first two fifths of the polypeptide chain folds to form the N-terminal helix, 5 of the 6 β strands and 2 of the 4 helices packed against the sheet; the following two fifths of the sequence forms the bulk of the random coil region as well as two helices, one packed around the 3-fold axis and one lying across the top of the sheet. The final one fifth of the chain forms the final β strand and two more helices packed against the β sheet, one of which is the C-terminal helix.

The apparent path of the polypeptide chain and the organisation of the secondary structure elements can now



Figure 6.11 a) A section of the 3.16 Å resolution electron density map perpendicular to the 'c' axis and in the plane of β strand 3.

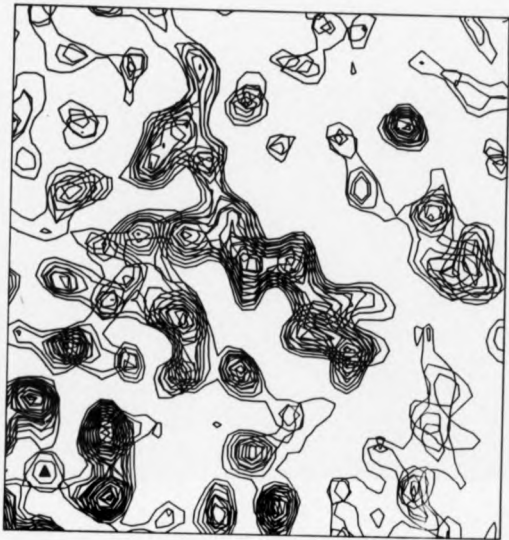


Figure 6.11 b) A similar section to that in a) , in this case lying approximately in the plane of β strand 5.

be described in more detail. The N-terminal helix, A, (see following paragraphs) lies in the x, y plane, close to $z = 0$, and forms part of the trimer-trimer interface within the hexamer. The chain then turns up into strand 4 of the β sheet and then turns into helix E which packs against strands 1 and 2. The major heavy atom site is located between this helix and strand 4, and the poorly defined electron density of this helix suggests some disturbance due to the heavy atom binding. Helix E turns into strand 1, the edge strand, out of which the chain follows an extended loop of random coil which provides another point of contact with the two-fold related trimer. This loop eventually turns into strand 2, the anti-parallel strand, followed by a sharp turn into strand 3. From this point the chain crosses to the other side of the sheet, passing close to the active site before turning into helix B and strand 5. A long region of random coil follows and includes a loop passing close to the 3-fold axis and to the active site of the adjacent 3-fold related subunit. The chain then forms helix D, lying across the top of the sheet, a turn into an extended conformation running anti-parallel to helix D, another loop past the active site, and then a turn into helix G, which comes into close contact with its symmetry equivalents around the 3-fold axis. Heavy atom site b is located along the exposed surface of this helix. The helix is followed by another loop past the active site, a turn into helix C and a sharp turn into strand 6 thereby completing the β sheet. The third heavy atom site, c, is a pocket formed between

helices B and C. The final portion of the sequence leaves strand 6, passing the active site into a long loop through the solvent region outside the active site pocket. The electron density along the loop is weak suggesting that this region may be somewhat flexible. Finally the chain turns into the C-terminal helix, F, which packs against the 8 sheet alongside helix E to complete the α domain.

The preliminary α carbon backbone tracing is shown in figure 6.12. The coordinates used to construct this tracing suffer in places from large errors - some adjacent α -carbons are separated by 2 Å and some by 5 Å - nevertheless, the drawing is a good representation. The region of random coil does contain what may perhaps be interpreted as a small

barrel lying between helices D and G. Figure 6.13 shows the relevant pieces of the chain in isolation.

With the amino acid sequence of PNP now available (see appendix I) use has been made of the various protein secondary structure prediction algorithms available. A package which combines the simplified output of some of these algorithms has been compiled by Dr. T. J. Greenhough (unpublished), including those of Chou and Fasman (1974, 1978, 1979), Cohen et.al. (1983), Garnier et.al. (1978), Lim (1974a, b), Burgess and Scheraga (1975) and Nagano (1977). These programs differ greatly in their method and in their relative strengths and weaknesses; however, it is not proposed that they be discussed here. In the case of the PNP study where an interpretable electron density map is available it is likely that the protein structure will be of more use in assessing



Figure 6.12 Stereoview looking down the 'c' axis of the initial α -carbon chain trace through the PNP monomer.

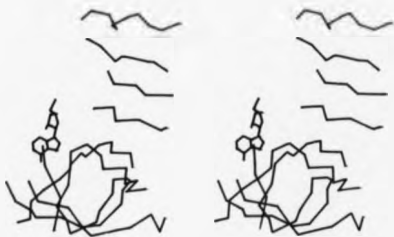


Figure 6.13 Stereoview of the α -carbon trace of the isolated β sheet and the extended chains which form a possible β barrel. The IFB model is also shown to indicate the region of the active site.

the prediction programs than the predictions will be in determining the structure. Nevertheless, the combined predictions have proved to be of some use, most notably in the location of the amino terminus of the polypeptide chain. All the programs predict (with varying degrees of confidence) the presence of an α -helix between residues 7 and 18. The amino acid sequence shows tyrosines at residues 5, 7 and 10, and a tryptophan at residue 16. Inspection of the electron density maps swiftly eliminated all the helices except for helix A which displays unambiguous side group density of the appropriate size for a tryptophan residue approximately three residues from its carboxyl terminus. In addition side group density appropriate for tyrosines is distinguishable in the expected positions at the amino end of the helix. The electron density beyond tyrosine 5 is then seen to peter out as is commonly observed with the often labile residues at either end of a polypeptide chain.

Beyond the N terminal helix the electron density is seen to turn upwards through histidine 23 for which there is appropriate side group density, and into strand 4 of the sheet. Most of the prediction programs indicate a β strand between residues 26 and 31. The largest peak in all the maps occurs at the carboxyl terminus of strand 4 and is likely to correspond to the cysteine residue at 31. It is also likely that this group is involved in the chemistry of the active site.

Efforts have been made to correlate the other strongly predicted helices with those observed in the electron density

map, with little success. However helix C has been matched with the residues between tyrosine 91 and histidine 104. Proline residues appear at 92 and 99 and, since proline is generally classed as a helix breaker, it is not surprising that the predictions fail to indicate the possibility of a helix in this region of the sequence. A proline residue in the middle of a helical region will tend to give rise to a bend of perhaps 20° in the otherwise straight helical axis. In the case of helix C a bending of the axis can be seen in the 3.16 \AA map. A stereo view of this helix with a helical model of residues 91 to 104 is shown in figure 6.14.

It should be pointed out that the analysis of the electron density maps is at an early stage and it is quite feasible that some of the specific interpretations described will be altered in due course.

6.5 Substrate binding studies

Preliminary studies of PNP-substrate complexes have begun with the calculation of an IFB difference Fourier map of 3.16 \AA . Coefficients in $F_{\text{IFB}} - F_{\text{NV}}$ were used and the phases were calculated as for the 3.16 \AA electron density map with the IFB contribution omitted. Two views of the difference density are shown in figure 6.15. The position of the IFB model relative to the crystallographic axes at 6 \AA was a good fit to the 3.16 \AA resolution density when these were first superimposed on the Multipicture system. To obtain the fit seen in fig. 6.15 the model has been rotated and translated by small amounts with the largest atomic shift

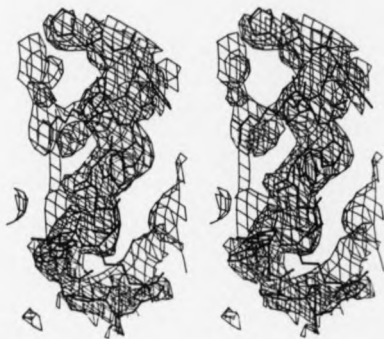


Figure 6.14 Stereoview of helix C superimposed on a helical model of residues Tyr 91 through to His 104.

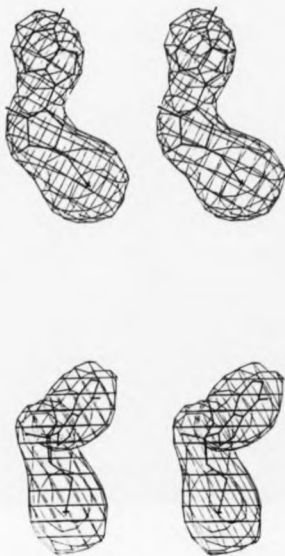


Figure 6.15 Two stereoviews of the IFB difference density at 3.16 Å resolution, with the fitted model superimposed.

being less than 1 \AA° .

Such studies will naturally be intensified once the residues in the active site region are identified and the high resolution data collected on the SRS wiggler beam line are processed.

CHAPTER SEVEN

CONCLUSION

Since the late 1970's SR has had a dramatic impact on macromolecular crystallography. Firstly, the high intensity has greatly reduced data collection times and enabled significantly more data to be recorded from individual crystal samples. Secondly, studies of weakly diffracting systems (small crystals or large unit cells) which were previously impossible or, at best, unattractive, can now be realistically entertained. Finally, the continuous spectrum of SR has led to the development of new approaches in structure analysis, including optimised anomalous scattering, Laue diffraction, and the polychromatic profile method. That every U.K. laboratory concerned with macromolecular crystallography has employed SR in one or more projects, confirms the significance of these developments.

Crystals of Purine Nucleoside Phosphorylase are weakly diffracting and radiation sensitive. Diffractometer measurements to 6.0 \AA resolution were possible, and indeed, the structure was initially solved and the subunit organisation revealed using these data. Using a rotating anode source, a single poor quality oscillation photograph is all that can be obtained from each sample. This approach would have led, at best, to an electron density map of poor quality at perhaps 4.0 \AA resolution. Using SR, the effective lifetime of each sample was improved by an order

of magnitude, a complete data set could be recorded using just 3 or 4 crystals, and an electron density map of good quality was produced at 3.16 Å resolution.

However, it is clear that the overall quality of the PNP data, as judged by its internal consistency, is only moderate and this must, in part, be due to the inherent nature of the crystals. The data quality would have been improved to some extent by the availability of sample absorption corrections and by the use of profile fitting during the film processing (of which there is some recent encouraging experience - see below).

Analysis of the heavy atom derivative data strongly suggests that the majority of the phase information was obtained from the EMTS derivative. Of the iodinated substrate analogues, IFB proved useful, but IG was discarded once doubts were raised concerning the presence of bound iodine. The PCMS proved the most problematical of the derivatives in that repeatable soaking conditions proved difficult to pinpoint (see table 5.1). Although the single, poorly occupied site observed with the SRS PCMS data corresponded to the EMTS major site, there was enough phase information to retain this derivative.

Although the interpretation of the PNP electron density map described in §6, awaits confirmation through model building of the sequence into the map, several aspects of the secondary and tertiary structure of the molecule are clear. The protein is a trimer with a single subunit in the crystallographic asymmetric unit, leaving the crystals with a relatively high solvent content of 78%. The subunit

structure falls into the class of parallel α/β proteins, having a six stranded β sheet flanked on each side by a pair of α -helices. Superficially, there is a close resemblance to the di-nucleotide binding domains observed in other structures. However, as the chain trace stands, there are sufficient differences in the topology of the secondary structural elements to suggest that PNP falls into a different subgroup of the parallel α/β category. Further examination of the completed model may reveal whether any evolutionary relationships between PNP and other proteins of this class are likely. Of additional interest would be a comparison of the PNP sequence against other protein sequences.

With the underlying aim of the PNP structure determination being one of rational drug design, knowledge of the geometric arrangement of atoms at the active site is obviously essential. While this information is not yet available, the general position of the active site is given by the IFB coordinates, and the difference Fourier maps calculated to date have suggested the likely orientation of the purine moiety. In addition, it has been suggested (S. E. Ealick, pers. comm.) that the large peak in the protein map thought to represent Cys 31 (see §6.4), is, in fact, a solvent sulphate ion. If so, this would presumably indicate the phosphate binding site. Once the protein structure is finalised and the diffraction data (collected on the wiggler station - see §4.6) from the protein-substrate complexes are available, there will be plenty of information with which to start a program of inhibitor design.

A consideration which arises when applying the principles of rational drug design to PNP inhibitors, is one of resolution. The PNP structure presented here is at 3.16 Å resolution and this represents the limit of the useful data currently available. Whether this will enable a sufficiently accurate and detailed model of the active site remains to be seen. Recent experiences with profile fitting techniques suggest that a useful number of significant data to 3.0 Å and beyond can be extracted from the best of the films already recorded (Greenhough and Suddath, (1986)). Some improvement in the resolution limit of the PNP data could probably be achieved by increasing the exposure time, recording the setting stills after one or two oscillation photographs, and then moving on to a fresh crystal.

The maximum resolution of the PNP diffraction data is undoubtedly limited by the very high solvent content of the crystals. It may well be worth the effort to search for an alternative crystal form in which the unit cell is more protein dense. If successful, molecular replacement methods could be applied to solve the new structure relatively quickly. In addition, there are various alternative sources of the protein and one of these may lead to a structure at higher resolution. Recently, crystals of a hexameric bacterial PNP have been prepared, and these diffract to about 2.6 Å resolution (S. E. Eslick & J. R. Helliwell, unpublished result).

In addition to the standard crystallographic studies

already under way, it would be worth considering the possibility of investigating the structure of intermediate states in PNP catalysis. The active site is easily accessible via the large solvent channels in the crystal and the binary complexes studied to date appear to be isomorphous with the native protein. The turnover rate of the reaction products would need to be matched to the data collection time and this may prove impossible to achieve. However, recent studies of phosphorylase b have shown this approach to be feasible with monochromatic data (see §1.3.2). In addition, Laue methods are currently being developed to improve the time resolution.

For ab initio determination of protein structure factor phase angles, at least one good quality isomorphous heavy atom derivative must be prepared and this can often prove very difficult or even impossible. Therefore, any new approach in which this requirement can be relaxed is potentially of great interest. The wavelength tunability of SR allows the possibility for optimised anomalous dispersion experiments in which phase angles are, in principle, uniquely determined by measurements on the native protein at two or more suitable, discrete wavelengths. The method is solely applicable to metalloproteins where the metal atom has a suitable K or L absorption edge, although a non-isomorphous, but well substituted, heavy atom derivative could be regarded as a native metalloprotein.

Close to an absorption edge, the magnitudes of f' and f'' vary rapidly with wavelength, particularly if a 'white line' feature is present. To make the best use of these

anomalous effects, spectral resolution is of the utmost importance. The Dy_2O_3 step scan experiment, described in §2.2.1, clearly demonstrates that resolutions better than 10^{-3} can be achieved, and that 'white line' effects are accessible with the singly bent, triangular monochromator as well as the double crystal system (station 9.6), both of which are in use at the SRS.

The formalism of Greenough & Helliwell, reviewed in §2.1, for beam geometry and spectral resolution, provides a completely general diffraction model for the single crystal monochromator configuration. This has led to an understanding of the left-right asymmetry in spot shape observed under non-Guinier experimental conditions, and to an appreciation of the relevant factors to be incorporated in the processing of SR diffraction data. Indeed, the need for this comprehensive formalism was, in part, originally stimulated by difficulties experienced in processing data measured at a non-Guinier setting at the EMBL, Hamburg.

By considering the relationship between a photon's energy and its direction in a non-Guinier situation, the concept of a polychromatic diffraction experiment arises. The theory shows that the average energy of each ray in the beam reflected from the monochromator, is correlated with the direction of that ray; and that consequently, each diffraction spot is drawn out into a streak, within which there is a linear relationship between X-ray energy and position along the streak axis. The theory also provides

a full description of the energy resolution at each point within such a diffraction streak in terms of the geometrical parameters of the experiment. By centring the wavelength spread on the absorption edge of a suitable anomalous scatterer within the crystal sample, a complete intensity profile across the edge can be recorded for each reflection. This polychromatic simultaneous profile method, therefore, represents an alternative to the discrete multi-wavelength approach and is, in principle, applicable to the same type of sample.

That the energy-direction correlation is preserved in the diffraction pattern, is the first and most obvious result of the Re experiment described in §2.2.2. Each spot in the 'on-edge' photographs clearly reveals the absorption 'white-line' at the Re edge. This feature would not be evident for a protein sample since the concentration of anomalous scatterer would be very much smaller. In the case of the Re experiment, the dominating absorption effect was removed by dividing the profile of interest by one (at similar θ angle) in which absorption was the only source of modulation. In the protein case, absorption streaks would have to be recorded under identical experimental conditions and at a series of inclination angles.

The fact that the 'divided profile' method used did reveal the expected large variation due to the Re f' effect, is highly encouraging for the future development of the technique. However, various questions remain to be answered before the method is successfully applied to

protein samples. The most notable of these is whether the information contained within the profiles can be extracted with sufficient accuracy to allow the structure factor phase to be well resolved. Current thought is centred on the way in which the profiles may be most usefully recorded to facilitate subsequent data reduction. In the future, it may well be that the aim of this experimental approach would be to provide an accurate subset of the phases as a starting point for the application of the direct methods currently under development for macromolecular structure determination.

REFERENCES

- Abrams R., Edmonds M. & Liberson L. (1965), Biochem. Biophys. Res. Comm., 20, 210
- Adams M.J., Blundell T.L., Dodson E.J., Dodson G.G., Vijayan M., Baker E.N., Harding M.M., Hodgkin D.C., Rimmer B. & Sheet S. (1969), Nature, 224, 491
- Agarwal K.C., Agarwal R.P. & Parks R.E.Jr. (1973), Federation Proc., 32, 581
- Agarwal K.C., Agarwal R.P., Stoeckler J.D. & Parks R.E.Jr. (1975), Biochemistry, 14, 79
- Agarwal R.C. (1978), Acta Cryst., A34, 791
- Agarwal R.P., Agarwal K.C. & Parks R.E.Jr. (1978), Methods in Enzymology, 51, 581
- Agarwal R.P. & Parks R.E.Jr. (1969), J. Biol. Chem., 244, 644
- Agarwal R.P. & Parks R.E.Jr. (1971), J. Biol. Chem., 246, 3763
- Agarwal R.P., Spector T. & Parks R.E.Jr. (1977), Biochem. Pharmac., 26, 359
- Alber T., Petsko G.A. & Tsernoglou D. (1976), Nature, 263, 297
- Amann A.J. (1978), Ciba Found. Symp., 68, 55
- Andrews S.J., Hails J. & Harding M.M. (1985), Unpublished work
- Arndt U.W., Champness J.N., Phizackerley R.P. & Wonacott A.J. (1973), J. Appl. Cryst., 6, 457
- Arndt U.W. & Gilmore D.J. (1979), J. Appl. Cryst., 12, 1
- Arndt U.W., Greenhough T.J., Helliwell J.R., Howard J.A.K., Rule S.A. & Thompson A.W. (1982), Nature, 298, 835
- Arndt U.W. & Thomas D.J. (1982), Nucl. Instr. Meth., 201, 21
- Arndt U.W. & Willis B.T.M. (1966), "Single Crystal Diffractometry", C.U.P.
- Arndt U.W. & Wonacott A.J. (1977), eds, "The Rotation Method in Crystallography", Elsevier/North-Holland Biomedical Press, Amsterdam
- Austel V. (1984), in "X-ray Crystallography and Drug Action", Horn A.S. & De Ranter C.J., eds, Clarendon Press, Ch.24

- Beddell A. (1984), in "X-ray Crystallography and Drug Action", Horn A.S. & De Ranter C.J., eds, Clarendon Press, Ch.10
- Blaschke G., Kraft H.P., Fickentscher K. & Kohler F. (1979), *Arzneim-Forsch*, 29, 1640
- Blow D.M. & Matthews B.V. (1973), *Acta Cryst.*, A29, 56
- Blow D.M. & Crick F.H.C. (1959), *Acta Cryst.*, 12, 794
- Blundell T.L. & Johnson L.N. (1976), "Protein Crystallography", Academic Press, London
- Bolin J.T., Filman D.J., Matthews D.A., Hamlin R.C. & Kraut J. (1982), *J. Biol. Chem.*, 257, 13650
- Bricogne G. (1984), *Acta Cryst.*, A40, 410
- Brittain R.T., Drev G.M. & Levy G.F. (1982), *Br. J. Pharmacology*, 77, 105
- Burgess A.N. & Scheraga H.A. (1975), *Proc. Natl. Acad. Sci. USA*, 72, 1221
- Burgess F.W., Stoeckler J.D. & Parks R.E.Jr., Unpublished results
- Camerman N. & Camerman A. (1984), in "X-ray Crystallography and Drug Action", Horn A.S. & De Ranter C.J., eds, Clarendon Press, Ch.13
- Cant J., Hellivell J.R., Merritt E., Phizackerley R.P. & Thompson A.V. (1983), Unpublished results
- Carlson J.D. & Fischer A.G. (1979), *Biochim. Biophys. Acta*, 571, 21
- Carr P.D. (1984), Ph.D. Thesis, Univ. of Keele
- Carson D.A., Kaye J., Matsumoto S., Seegmiller J.E. & Thompson L. (1979), *Proc. Natl. Acad. Sci. USA*, 76, 2430
- Chan T. (1978), *Cell*, 14, 523
- Chou P.Y. & Fasman G.D. (1974), *Biochemistry*, 13, 222
- Chou P.Y. & Fasman G.D. (1978), *Ann. Rev. Biochem.*, 47, 251
- Chou P.Y. & Fasman G.D. (1979), *Biophys. J.*, 26, 367
- Chothia C. (1973), *J. Mol. Biol.*, 75, 295
- Clark G., Gani S., Hellivell J.R. & Tanner B.K. (1981), Unpublished results
- Cocco L., Roth B., Temple C., Montgomery J.A., London R.E. & Blakley R.L. (1983), *Arch. Biochem. Biophys.*, 226, 567

- Cohen A., Doyle D., Martin D.W.Jr. & Ammann A.J. (1976), *New England J. Medicine*, 295, 1449
- Cohen F.E., Abarbanel R.M., Kuntz I.D. & Fletterick R.J. (1983), *Biochemistry*, 22, 4894
- Cook V.J., Ealick S.E., Bugg C.E., Stoeckler J.D. & Parks R.E.Jr. (1981), *J. Biol. Chem.*, 256, 4079
- Cori C.F., Colovick S.P. & Cori G.T. (1937), *J. Biol. Chem.*, 121, 465
- Covan M.J., Cashman D. & Ammann A.J. (1981), *Biochem. Pharmac.*, 30, 2651
- Crate J., Cook V.J., Ealick S.E. & Bugg C.E. (1980), Unpublished results
- Crick F.H.C. & Magdoff B. (1956), *Acta Cryst.*, 9, 901
- Cromer D.T. (1983), *J. Appl. Cryst.*, 16, 473
- Deisenhofer J. & Steigemann W. (1975), *Acta Cryst.*, B31, 238
- Diamond R. (1971), *Acta Cryst.*, A25, 436
- Diamond R. (1974), *J. Mol. Biol.*, 82, 371
- Dickerson R.E., Kendrew J.C. & Standberg B.E. (1961), *Acta Cryst.*, 14, 1188
- Dodson E.J., Evans P.R. & French S. (1975), in "Anomalous Scattering", eds, Ramasesham S. & Abrahams S.C., Munksgaard, Copenhagen
- Dodson E.J. & Vijayan M. (1971), *Acta Cryst.*, B27, 2402
- Dosch H.-M., Mansour A., Cohen A., Shore A. & Gelfand E.V. (1980), *Nature*, 285, 494
- Dowd D.J., Chen S.-F., Stoeckler J.D. & Parks R.E.Jr. (1983), Unpublished results
- Duax W.L., Griffin J.F. & Rohrer D.C. (1981), *J. Am. Chem. Soc.*, 103, 6705
- Ealick S.E., Greenhough T.J., Babu Y.S., Cook V.J., Bugg C.E., Rule S.A., Habash H. & Hellivell J.R. (1984), *Acta Cryst.*, A40, C30
- Edwards Y.H., Edwards P.A. & Hopkinson D.A. (1973), *F.E.B.S. letters*, 32, 235
- Edwards Y.H., Hopkinson D.A. & Harris H. (1971), *Ann. Hum. Genet. Lond.*, 24, 395

- Eisenberger P., Newsam J.M., Leonevitz M.E. & Vaughan D.E.W. (1984), *Nature*, 309, 45
- Elion G.B., Furman P.A., Fyle J.A., de Miranda P., Beauchamp L. & Schaeffer H.J. (1977), *Proc. Natl. Acad. Sci. USA*, 74, 5716
- Evald P.P. (1921), *Z. Kristolleg. Miner.*, 56, 129
- Fox G.C. & Holmes K.C. (1966), *Acta Cryst.*, 20, 886
- Friedkin M. (1950), *J. Biol. Chem.*, 184, 449
- Friedkin M. & Kalckar H.M. (1966), *J. Biol. Chem.*, 184, 437
- Garnier J., Osguthorpe D.J. & Robson B. (1978), *J. Mol. Biol.*, 120, 97
- Gelfand E.W., Dosch H.-M., Biggar W.D. & Fox I.H. (1978), *J. Clin. Invest.*, 61, 1071
- Changes G. & Rheem G.B. (1979), *J. Biol. Chem.*, 254, 4273
- Giblett E.R., Ammann A.J., Vara D.V., Sandman R. & Diamond L.K. (1975), *Lancet*, 1, 1010
- Giblett E.R., Anderson J.E., Cohen F., Pollara B. & Meavissen H.J. (1972), *Lancet*, 2, 1067
- Goddard J.M., Caput D., Williams S.R. & Martin D.W.Jr. (1983), *Proc. Natl. Acad. Sci. USA*, 80, 4281
- Green D.V., Ingram V.M. & Perutz M.F. (1954), *Proc. Royal Society, A*, 225, 287
- Greenhough T.J. & Hellivell J.R. (1982a), *J. Appl. Cryst.*, 15, 338
- Greenhough T.J. & Hellivell J.R. (1982b), *J. Appl. Cryst.*, 15, 493
- Greenhough T.J., Hellivell J.R. & Rule S.A. (1983), *J. Appl. Cryst.*, 16, 242
- Greenhough T.J. & Suddath F.L. (1986), *J. Appl. Cryst.*, 19, 400
- Gudas L.J., Ullman B., Cohen A. & Martin D.W.Jr. (1978a), *Cell*, 14, 531
- Gudas L.J., Zannis V.I., Clift S.M., Ammann A.J., Staal G.E.J. & Martin D.W.Jr. (1978b), *J. Biol. Chem.*, 253, 8916
- Hajdu J., Acharya K.R., Stuart D.I., McLaughlin P.J., Barford D., Oikonomako N.G., Klein H. & Johnson L.N. (1987), *EMBO J.*, 6, 539
- Hansen J.C., Watenpaugh K.D., Sieker L. & Jensen L.H. (1979), *Acta Cryst.*, A35, 616

- Harding M.M. (1962), D. Phil. Thesis, Univ. of Oxford
- Hellivell J.R. (1981), Unpublished results
- Hellivell J.R. (1984), Reports Prog. Phys., 47, 1403
- Hellivell J.R. et. al. (1985), in preparation for submission to J. Appl. Cryst.
- Hellivell J.R., Greenhough T.J., Carr P.D., Rule S.A., Moore P.R., Thompson A.V. & Vorgan J.S. (1982a), J. Phys. E, 15, 1363
- Hellivell J.R., Hughes G., Pryzbylski M.M., Ridley P.A., Sumner I., Bateman J.E., Connelly J.F. & Stephenson R. (1982b), Nucl. Instr. Meth., 201, 175
- Hellivell J.R., Moore P.R., Papiz M.Z. & Smith J.M.A. (1984), J. Appl. Cryst., 17, 417
- Hellivell J.R., Papiz M.Z., Glover I.D., Habash J., Thompson A.V., Moore P.R., Harris N., Croft D. & Pantos E. (1986), Nucl. Instr. Meth., A246, 617
- Hendrickson V.A. & Teeter M.M. (1981), Nature, 290, 107
- Hendrix J., Koch M. & Bordas J. (1979), J. Appl. Cryst., 12, 467
- Herriott J.R., Sieker L.H., Jensen L.H. & Lovenberg W. (1970), J. Mol. Biol., 50, 391
- Hershfield M.S. (1981), J. Clin. Invest., 67, 696
- Hershfield M.S., Kredick N.M., Ownby D.R. & Buckley R. (1979), J. Clin. Invest., 63, 807
- Hitchings G.H. & Smith S.L. (1980), Adv. Enzyme Regulation, 18, 349
- Hoppe V. & Jakubovski U. (1975), in "Anomalous Scattering", eds, Ramasesham S. & Abrahams S.C., Munksgaard, Copenhagen
- Howard J.A.K., Mead K.A. & Spencer J.L. (1983), Acta Cryst., C39, 555
- Ikazawa Z., Nishino T., Murakami K. & Tsushima K. (1978), Comp. Biochem. Physiol., 60B, 111
- Irvin M.J., Nyborg J., Reid B.R. & Blow D.M. (1976), J. Mol. Biol. 105, 577
- Jack A. & Levitt M. (1978), Acta Cryst., A34, 931
- Jackson J.D. (1975), "Classical Electrodynamics", Wiley, New York
- Jordan F. & Wu A. (1978a), Arch. Biochem. Biophys., 190, 699

- Jordan F. & Wu A. (1978b), J. Med. Chem., 21, 877
- Kahn R., Fourme R., Bosshard R., Chiadmi M., Risler J.L., Dideberg O. & Véry J.P. (1985), F.E.B.S. letters, 179, 133
- Kahn R., Fourme R., Gadet A., Janin J., Dumas C. & Andre D. (1982), J. Appl. Cryst., 15, 330
- Kalata K. (1982), Nucl. Instr. Meth., 201, 35
- Kalckar H.M. (1945a), Federation Proc., 4, 248
- Kalckar H.M. (1945b), J. Biol. Chem., 158, 723
- Kalckar H.M. (1947), J. Biol. Chem., 167, 477
- Kamatani N. & Carson D.A. (1980), Cancer Res., 40, 4178
- Kamzers I.S., Mitchell B.S., Dadonna P.E., Wotring L.L., Townsend L.B. & Kelly V.N. (1981), Science, 214, 1137
- Kendrew J.C., Bodo G., Dintzis H., Parrish A.G., Wycoff H. & Phillips D.C. (1958), Nature, 181, 662
- Kim B.K., Chu S.-H. & Parks R.E.Jr. (1968), J. Biol. Chem., 243, 1763
- Klein W. (1935), Z. Physiol. Chem., 231, 125
- Klotz I.W., Darnall D.V. & Langerman N.R. (1975), in "The Proteins" Vol 1, eds, Neurath H. & Hill R.L., Academic Press, New York
- Kohra K., Ando M., Matsushita T. & Hashizume H. (1978), Nucl. Instr. Meth., 152, 161
- Konnert J.H. & Hendrickson V.A. (1980), Acta Cryst., A36, 344
- Koyama G. & Umezawa H. (1965), J. Antibiotics (Tokyo), 18A, 175
- Kraut J., Sieker L.C., High D.F. & Freer S.T. (1962), Proc. Natl. Acad. Sci. USA, 48, 1417
- Krenitsky T.A. (1967), Mol. Pharmac., 3, 526
- Krenitsky T.A., Elion G.B., Henderson A.M. & Hitchings G.H. (1968), J. Biol. Chem., 243, 2876
- Krenitsky T.A., Elion G.B., Strelitz R.A. & Hitchings G.H. (1967), J. Biol. Chem., 242, 2675
- Kunin C.M. (1970), in "Medicinal Chemistry" vol. 1 p246,ed, Burger A., Wiley, New York
- Lemonnier M., Fourme R., Rousseaux F. & Kahn R. (1978), Nucl. Instr. Meth., 152, 173

- Lepage G.A., Junga I.G. & Bowman B. (1964), *Cancer Res.*, 24, 835
- Lerman L.S. (1961), *J. Mol. Biol.*, 3, 18
- Lewis A.S. (1978), *Arch. Biochem. Biophys.*, 190, 662
- Lewis A.S. & Glants N.D. (1976a), *J. Biol. Chem.*, 251, 407
- Lewis A.S. & Glants N.D. (1976b), *Biochemistry*, 15, 4451
- Lewis A.S. & Lovry B.A. (1979), *J. Biol. Chem.*, 254, 9927
- Lim V.I. (1974a), *J. Mol. Biol.*, 88, 857
- Lim V.I. (1974b), *J. Mol. Biol.*, 88, 872
- Lineveaver H. & Burk D. (1934), *J. Amer. Chem. Soc.*, 56, 638
- Martin F.V. & Cacak R.K. (1976), *J. Phys. E*, 9, 662
- Matelik G. (1982), in "Uses of Synchrotron Radiation In Biology", ed, Stuhrman B.B., Academic Press, London
- Matthews B.V. (1966), *Acta Cryst.*, 20, 230
- Matthews B.V. (1968), *J. Mol. Biol.*, 33, 491
- Matthews D.A., Alden R.A., Bolin J.T., Freer S.T., Hamlin R.C., Xuong N., Kraut J., Poe M., Williams M. & Hoogsteen K.V. (1977), *Science*, 197, 452
- McRoberts J.A. & Martin D.V.Jr. (1980), *J. Biol. Chem.*, 255, 5605
- Milman G., Anton D.L. & Veber J.L. (1976), *Biochemistry*, 15, 4967
- Moffat K., Szebenyi D. & Bilderback D. (1984), *Science*, 223, 1423
- Montgomery J.A., Schabel P.M.Jr. & Skipper H.E. (1962), *Cancer Res.*, 22, 504
- Moore E.C. & Hurlbert R.B. (1966), *J. Biol. Chem.*, 241, 4802
- Morimoto H. & Uyeda R. (1963), *Acta Cryst.*, 16, 1107
- Moyer T.P. & Fischer A.G. (1976), *Arch. Biochem. Biophys.*, 174, 622
- Muirhead H., Cox J.N., Mazzarella L. & Perutz M.F. (1967), *J. Mol. Biol.*, 28, 117
- Murakami K. & Tsushima K. (1975), *Biochim. Biophys. Acta.*, 384, 390
- Nagano K. (1977), *J. Mol. Biol.*, 109, 251

- Narayan R. & Rameshan S. (1981), *Acta Cryst.*, A37, 636
- Navia M. & Sigler P. (1974), *Acta Cryst.*, A30, 706
- Neidle S. (1973), *Acta Cryst.*, B29, 2645
- Neidle S. (1984), in "X-ray Crystallography and Drug Action", Horn A.S. & De Ranter C.J., eds, Clarendon Press, Ch.8
- Neidle S. & Berman H.M. (1983), *Prog. Biophys. Mol. Biol.*, 41, 43
- Nishida Y., Kamatani N., Tanimoto K. & Akaoki I. (1979), *Agents and Actions*, 9, 549
- North A.C.T. (1965), *Acta Cryst.*, 18, 212
- North A.C.T., Phillips D.C. & Matthews F.S. (1968), *Acta Cryst.*, A24, 351
- Osbourne V.R.A. (1980), *J. Biol. Chem.*, 255, 7089
- Osbourne V.R.A., Sullivan J.L. & Scott C.R. (1980), *Immun. Comm.*, 2, 257
- Parks R.E.Jr. & Agarwal R.P. (1972), in "The Enzymes" Vol 3, ed., Boyer F.D., Academic Press, New York
- Parks R.E.Jr., Crabtree G.V., Kong C.M., Agarwal R.P., Agarwal K.C. & Scholar E.H. (1975), *N.Y. Acad. Sci.*, 255, 412
- Parks R.E.Jr., Savarese T.M. & Chu S.-H. (1982), in "New Approaches to the Design of Antineoplastic Agents", eds, Bardos T.J. & Kalman T.I., Elsevier Biomedical, New York
- Parks R.E.Jr., Stoeckler J.D., Cambor C., Savarese T.M., Crabtree G.V. & Chu S.-H. (1981), in "Molecular Actions and Targets for Cancer Chemotherapeutic Agents", eds, Sartorelli A., Lazo J.B. & Bertino J.R., Academic Press, New York
- Pearl L.H. & Honegger A. (1983), *J. Mol. Graphics*, 1, 9
- Perutz M.F. (1985), *New Scientist*, 1460, 12
- Phillips J.C. (1978), Ph.D. Thesis, Univ. of Stanford
- Ricciuti F. & Ruddle F.H. (1973), *Nat. New Biol.*, 241, 180
- Richardson J.S. (1981), *Advances in Protein Chemistry*, 34, 167
- Rossmann M.G. (1972), "The Molecular Replacement Method", Gordon & Breach, New York
- Rossmann M.G. & Erikson J. (1983), *J. Appl. Cryst.*, 16, 629
- Savage B. & Spencer N. (1977), *Biochem. J.*, 167, 703

- Schaeffer H.J., Beauchamp L., de Miranda P., Elion G.B., Bauer D.J. & Collins P. (1978), *Nature*, 272, 583
- Schmid M.F., Weaver L.H., Holmes M.A., Grutter M.G., Ohlendorf D.H., Renolds R.A., Remington S.J. & Matthews B.V. (1981), *Acta Cryst.*, A37, 701
- Siav M.F.E., Mitchell B.S., Koller C.A., Coleman M.S. & Hutton J.J. (1980), *Proc. Natl. Acad. Sci. USA*, 77, 6157
- Siegenbeek van Heukelom L.H., Akkerman J.W.N., Staal C.E.J., DeBrayn C., Stoop J.V., Zegers B.J.M., DeBree F.K. & Vadmán S.K. (1977), *Clin. Chim. Acta*, 74, 271
- Sielecki A.R., Hendrickson V.A., Broughton C.G., Delbaure L., Brayer G.D. & James M.N.G. (1979), *J. Mol. Biol.*, 134, 781
- Simmonds H.A., Levinsky R.J., Perrett D. & Webster D.R. (1982), *Biochem. Pharmac.*, 31, 947
- Smith J.L., Hendrickson V.A. & Adison A.V. (1983), *Nature*, 303, 86
- Srinivasen R. (1961), *Acta Cryst.*, 14, 607
- Steitz T.A. (1968), *Acta Cryst.*, B24, 504
- Stoeckler J.D., Agarwal R.P., Agarwal K.C. & Parks R.E.Jr. (1978b), *Methods in Enzymology*, 51, 530
- Stoeckler J.D., Agarwal R.P., Agarwal K.C. & Schmid K. (1978a), *Biochemistry*, 17, 278
- Stoeckler J.D., Bell C., Parks R.E.Jr., Chu C.K., Fox J. & Ikehara M. (1982b), *Biochem. Pharmac.*, 31, 1723
- Stoeckler J.D., Cambor C., Burgess F., Erban S.B. & Parks R.E.Jr. (1980), *Pharmacologist*, 22, 99
- Stoeckler J.D., Cambor C., Kahns V., Chu S.-H. & Parks R.E.Jr. (1982a), *Biochem. Pharmac.*, 31, 163
- Stoeckler J.D., Klein R.S. & Parks R.E.Jr. (1983), Unpublished results
- Stoeckler J.D., Stein J.M. & Parks R.E.Jr. (1984), Unpublished results
- Tax W.J.M. & Veerkamp J.H. (1978), *Comp. Biochem. Physiol.*, 61B, 439
- Templeton D.H. & Templeton L.K. (1985), *Acta Cryst.*, A41, 365
- Templeton L.K., Templeton D.H. & Phizackerley R.P. (1980), *J. Am. Chem. Soc.*, 102, 1185

- Tickle I.J. (1982), "Molecular Interactive Display And Selection", Birkbeck College, London
- Townsend L.B., Cline B.L., Panzica R.P., Fagerness P.E., Roti L.W., Stoeckler J.D., Crabtree G.W. & Parks R.E.Jr. (1978), in "Lectures in Heterocyclic Chemistry" Vol 4, eds, Castle R.N. & Lalezari I., S79 Hetero Corporation USA
- Tsubio K.K. & Hudson P.B. (1957), J. Biol. Chem., 224, 879
- Turner B.M., Fisher R.A. & Harris H. (1971), Eur. J. Biochem., 24, 288
- Turner B.M., Fisher R.A. & Harris H. (1974), Clin. Chim. Acta, 50, 85
- Tuttle J.V. & Krenitsky T.A. (1984), J. Biol. Chem., 259, 4065
- Ullman B., Gudas L.J., Clift S.M. & Martin D.W.Jr. (1979), Proc. Natl. Acad. Sci. USA, 76, 1074
- Valentine W.N., Fink K., Paglia D.E., Harris S.R. & Adams W.S. (1974), J. Clin. Invest., 54, 866
- White P.F., Ham J., Way W.L. & Trevor A.J. (1980), Anesthesiology, 52, 231
- Wiginton D.A., Coleman M.S. & Hutton T.J. (1980), J. Biol. Chem., 255, 6663
- Witkowski J.T., Robins R.K., Khore G.P. & Sidwell R.W. (1971), J. Med. Chem., 16, 935
- Willemot J., Martineau R., DesRosiers C., Kelly S., Letourneau J. & Lalanne M. (1979), Life Sci., 25, 2515
- Williams S.R., Goddard J.M. & Martin D.W.Jr. (1984), Nucleic Acids Res., 12, 5779
- Willis R.C., Robins R.K. & Seegmiller J.E. (1980), Molec. Pharmac., 18, 287
- Wilson K.S. (1974), D. Phil. Thesis, Univ. of Oxford
- Wilson K.S., Stura E.A., Wild D.L., Todd R.J., Stuart D.I., Babu Y.S., Jenkins J.A., Standing T.S., Johnson L.N., Fourme R., Kahn R., Gadet A., Bartels K.S. & Bartunik H.D. (1983), J. Appl. Cryst., 16, 28
- Winick H. & Bienenstock A. (1978), Ann. Rev. Nucl. Part. Sci., 28, 33
- Vonacott A.J. (1977), in "The Rotation Method in Crystallography", eds, Arndt U.W. & Vonacott A.J., Elsevier/North-Holland Biomedical Press, Amsterdam

- Woolfson M.M. (1970), "An Introduction to X-ray Crystallography", C.U.P.
- Woolfson M.M. (1971), Reports Prog. Phys., 34, 369
- Xuong N.H., Freer S.T., Hamlin R., Nielson C. & Vernon V. (1978), Acta Cryst., A34, 289
- Xuong N.H., Sullivan D., Nielson C. & Hamlin R. (1985), Acta Cryst., B41, 267
- Zannis V.I., Doyle D. & Martin D.W.Jr. (1978), J. Biol. Chem., 253, 504
- Zannis V.I., Gudas L.J. & Martin D.W.Jr. (1979), Biochem. Genet., 17, 621
- Zimmerman T.P., Gersten N. B., Ross A.F. & Miech R.P. (1971), Can. J. Biochem., 49, 1050

APPENDIX I

The amino acid sequence given below (using the standard 3 letter code), has been derived from the nucleotide sequence of a cDNA clone which contains the entire coding region of the human PNP message (Williams et. al. (1984)). Active PNP protein from this cDNA clone has been successfully expressed in PNP-deficient *E. coli* cells by Goddard et. al. (1983).

```

1           10
met glu asn gly tyr thr tyr glu asp tyr lys asn thr ala glu trp leu leu

20           30
ser his thr lys his arg pro gln val ala ile ile cys gly ser gly leu gly

40           50
gly leu thr asp lys leu thr gln ala gln ile phe asp tyr ser glu ile pro

60           70
asn phe pro arg ser thr val pro gly his ala gly arg leu val phe gly phe

80           90
leu asn gly arg ala cys val met met gln gly arg phe his met tyr glu gly

100
tyr pro leu trp lys val thr phe pro val arg val phe his leu leu gly val

110           120
asp thr leu val val thr asn ala ala gly gly leu asn pro lys phe glu val

130           140
gly asp ile met leu ile arg asp his ile asn leu pro gly phe ser gly gln

150           160
asn pro leu arg gly pro asn asp glu arg phe gly asp arg phe pro ala met

170           180
ser asp ala tyr asp arg thr met arg gln arg ala leu ser thr trp lys gln

190
met gly glu gln arg glu leu gln glu gly thr tyr val met val ala gly pro

200           210
ser phe glu thr val ala glu cys arg val leu gln lys leu gly ala asp ala

220           230
val gly met ser thr val pro glu val ile val ala arg his cys gly leu arg

240           250
val phe gly phe ser leu ile thr asn lys val ile met asp tyr glu ser leu

260           270
gly lys ala asn his glu glu val leu ala ala gly lys gln ala ala gln lys

280
leu glu gln phe val ser ile leu met ala ser ile pro leu pro asp lys ala

289
met

```


Appendix 2

Publications

1. Central data collection facility for protein crystallography, small angle diffraction and scattering at the Daresbury Laboratory Synchrotron Radiation Source (SRS), England.

J. R. Helliwell, T. J. Greenhough, P. D. Carr, S. A. Rule,
P. R. Moore, A. W. Thompson and J. S. Vorgan.

J. Phys. E: Sci. Instrum., Vol 15, 1982, pp. 1363-1372

2. Optimized anomalous dispersion in crystallography: a synchrotron X-ray polychromatic simultaneous profile method.

U. V. Arndt, T. J. Greenhough, J. R. Helliwell, J. A. K. Howard,
S. A. Rule and A. W. Thompson.

Nature, Vol. 298, No. 5877, 1982, pp. 835-838.

3. Oscillation camera data processing. 3. General diffraction spot size, shape and energy profile: formalism for polychromatic diffraction experiments with monochromatized synchrotron X-radiation from a singly bent triangular monochromator.

T. J. Greenhough, J. R. Helliwell and S. A. Rule.

J. Appl. Cryst., Vol 16, 1983, pp. 242-250.

IMAGING SERVICES NORTH

Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
www.bl.uk

**PUBLISHED PAPERS NOT SCANNED
AT THE REQUEST OF THE
UNIVERSITY.**